



# A CLASS OF GROWTH MODELS RESCALING TO KPZ

MARTIN HAIRER<sup>1</sup> and JEREMY QUASTEL<sup>2</sup>

<sup>1</sup> Department of Mathematics, Imperial College London, London SW7 2AZ, UK;  
email: m.hairer@imperial.ac.uk

<sup>2</sup> Department of Mathematics, University of Toronto, M5S 1L2, Canada;  
email: quastel@math.toronto.edu

Received 20 April 2017; accepted 24 May 2018

## Abstract

We consider a large class of  $1 + 1$ -dimensional continuous interface growth models and we show that, in both the weakly asymmetric and the intermediate disorder regimes, these models converge to Hopf–Cole solutions to the KPZ equation.

2010 Mathematics Subject Classification: 60H15 (primary); 35R60, 82C28 (secondary)

## CONTENTS

1. Introduction	2
1.1 Intermediate disorder scaling	6
1.2 Weakly asymmetric scaling	9
1.3 Possible generalizations	11
1.4 Standing assumptions and terminology	13
2. Methodology	14
3. Construction of the regularity structure	16
3.1 The collection of symbols	16
3.2 Structure group	19
3.3 The extended regularity structure $\mathcal{T}_{\text{ex}}$	21
3.4 Admissible models	23
3.5 Definition of $\mathcal{D}^\gamma$	26

3.6	Canonical lift to $\mathcal{T}_{\text{ex}}$	27
3.7	Multiplication by $\varepsilon^k$	29
4.	Abstract solution map	31
4.1	Dealing with irregular initial conditions	34
4.2	Properties of the spaces $\mathcal{D}_{\varepsilon}^{\gamma,\eta}$	37
4.3	Operations in $\mathcal{D}_{\varepsilon}^{\gamma,\eta}$	40
4.4	Picard iteration and convergence	43
5.	Renormalization	47
5.1	Renormalization of the average speed	47
5.2	Wick renormalization	48
5.3	Renormalized equations	53
6.	Convergence of the models	56
6.1	A convergence criterion	57
6.2	Proof of Theorem 6.1	60
6.3	Behaviour of the renormalization constants	70
7.	Main convergence results	80
7.1	Weak asymmetry regime	80
7.2	Intermediate disorder regime	83
	Appendix A. A bound on generalized convolutions	85
A.1	Decomposition	89
A.2	Multiscale clustering	92
A.3	General form of the bound	96
A.4	Naive bound	97
A.5	Improved bound	98
A.6	Putting everything together	101
	Appendix B. Notes on renormalization	102
	Appendix C. Symbolic index	109
	References	110

## 1. Introduction

The Kardar–Parisi–Zhang equation is formally given by

$$\partial_t h^{(\lambda)} = \partial_x^2 h^{(\lambda)} + \lambda (\partial_x h^{(\lambda)})^2 + \xi, \quad (1.1)$$

where  $\xi$  denotes space–time white noise and  $\lambda \in \mathbf{R}$  is a parameter describing the strength of its ‘asymmetry’. Equation (1.1) should be interpreted either via the

Hopf–Cole transform [BG97] as

$$h_{\text{HC}}^{(\lambda)} \stackrel{\text{def}}{=} \frac{1}{\lambda} \log Z^{(\lambda)} \quad (1.2)$$

where  $Z^{(\lambda)}$  is the continuous [Wal86], strictly positive [Mue91] Itô solution of the multiplicative stochastic heat equation

$$dZ^{(\lambda)} = \partial_x^2 Z^{(\lambda)} + \lambda Z^{(\lambda)} dW, \quad Z^{(\lambda)}(0) = Z_0, \quad (1.3)$$

where  $Z_0 = \exp(\lambda h_0)$  with  $W$  an  $L^2$ -cylindrical Wiener process,  $\langle W_t - W_s, \varphi \rangle = \xi(\varphi \otimes \mathbf{1}_{[s,t]})$  or equivalently by using the theory exposed in [Hai13, Hai14]. It has been conjectured (see [BPRS93, BG97, GJ14] for a number of results in this direction) that the KPZ equation has a ‘universal’ character in the sense that any one-dimensional model of surface growth should converge to it provided that it has the following features:

- There is a microscopic smoothing mechanism.
- The system has microscopic fluctuations with short-range correlations.
- The system has some ‘lateral growth’ mechanism in the sense that the growth speed depends in a nontrivial way on the slope.
- At the microscopic scale, the strengths of the growth and fluctuation mechanisms are well separated: either the growth mechanism dominates (intermediate disorder) or the fluctuations dominate (weak asymmetry).

Only some progress has been made towards a rigorous mathematical understanding of this claim. The only discrete microscopic models for which convergence to the KPZ equation has been established rigorously in general are the height function of asymmetric exclusion processes in the weakly asymmetric limit [BG97], [ACQ11], [DT13], qTASEP [BC14, CT15] and the free energy of directed random polymers in the intermediate disorder regime [AKQ10], [MFQR]. In [GJ14] it was shown that a wide class of asymmetric particle models with product invariant measures converge to *energy solutions* of the KPZ equation when started in equilibrium. A slightly stronger version of these equilibrium energy solutions were shown to be unique in [GP15]. In the continuous setting [FQ14] consider the KPZ equation with nonlinearity smoothed out so that a smoothed out Brownian motion is invariant, and show, again, that in equilibrium it converges to KPZ. In all these cases, including the last two, the proof goes through the Hopf–Cole transformation, and relies on the result satisfying a manageable version of (1.3). This is avoided in

the regularity structures approach [Hai13, Hai14] which, in principle, allows for many different types of regularization of the quadratic KPZ equation or stochastic heat equation [HP15]. At the present time it is however restricted to finite volume.

Substantial progress has also been made recently in the understanding of the conjectured long time scaling limit of the KPZ equation itself, which is expected to be the scaling limit for this whole class of microscopic interface growth models [Spo91, BQS11, ACQ11, BC14]. Note that the type of well-posedness and approximation results considered here, or in [Hai14], even when they are global, do not have much to say about large time, which presently can only be probed through exact calculations.

In this article, we consider continuous growth models of the type

$$\partial_t h = \partial_x^2 h + \varepsilon F(\partial_x h) + \delta \eta, \quad (1.4)$$

where  $F$  is an even function, which we will often take to be a polynomial, modelling the growth mechanism,  $\eta$  is a *smooth* space–time Gaussian process modelling the microscopic fluctuations, and  $\varepsilon, \delta$  are two parameters. The two regimes alluded to earlier correspond to  $\varepsilon \approx 1$  and  $\delta \ll 1$  (intermediate disorder), as well as  $\varepsilon \ll 1$  and  $\delta \approx 1$  (weak asymmetry). It is important to note that these two regimes are *not* equivalent, that is, it is not possible to turn one regime into the other by a simple change of variables. What is usually done is to formally expand

$$F(s) = F(0) + F'(0)s + \frac{1}{2}F''(0)s^2 + \dots$$

The first two terms in the expansion can be removed by simple height and spatial shifts and one argues that the model is then approximated by the quadratic KPZ equation (1.1) with  $\lambda = \frac{1}{2}F''(0)$  [HHZ95, KS91].

Our main result is that for a wide class of nonlinearities  $F$  and correlation functions for  $\eta$ , the appropriate rescaling of (1.4) (as a function of the small parameter  $\varepsilon$  or  $\delta$  depending on the regime considered) converges to the KPZ equation (1.1) for a suitable value of the parameter  $\lambda$ . While this result is to some extent expected in view of the above discussion, the precise analysis uncovers some surprising facts:

- In the weakly asymmetric regime, the value  $\lambda$  obtained for the limiting equation is *not* the one that one would guess by formally rescaling the equation and neglecting all terms with a positive power of the small parameter. In particular, one generically has  $\lambda \neq 0$  even if the polynomial  $F$  has no quadratic term.

- In the intermediate disorder regime, if we consider  $F$  with  $F''(0) = 0$  but  $F^{(4)}(0) \neq 0$  (say) then, as expected, the limit obtained under the ‘naive rescaling’ is given by the additive (linear) stochastic heat equation (1.1) with  $\lambda = 0$ . However, by considering larger scales, one again recovers the KPZ equation with a nontrivial  $\lambda$ !

To understand the need for the separation of scales, let us consider the problem of trying to make sense of (1.4) with  $\varepsilon = \delta = 1$ , when  $\eta$  is space–time white noise. The natural approach is to replace  $\eta$  by an approximate white noise  $\xi^{(\gamma)}$  which is smooth on some small scale  $\gamma > 0$  and attempt to identify a limit of

$$\partial_t h_\gamma = \partial_x^2 h_\gamma + F(\partial_x h_\gamma) + \xi^{(\gamma)}.$$

In the KPZ case,  $F(u) = u^2$ , the nonlinear term does indeed converge to a nontrivial field, at the simplest level in the sense of convergence of the space–time covariance, after renormalization by subtraction of a diverging constant. On the other hand, if one takes a higher order nonlinearity such as  $F(v) = v^4$ , the renormalization by constants cannot help: The space–time covariance of the nonlinear field simply diverges as  $\gamma^{-2}$ . A possible route might be to renormalize by subtracting quadratic terms. For example, one could try to take a limit of

$$\partial_t h_\gamma = \partial_x^2 h_\gamma + [(\partial_x h_\gamma)^4 - c_{2,\gamma}(\partial_x h_\gamma)^2 + c_{1,\gamma}] + \xi^{(\gamma)}, \quad (1.5)$$

with precisely chosen  $c_{1,\gamma}$  and  $c_{2,\gamma}$ . The model is supercritical, and on large scales one expects such systems to be diffusive, that is, to exhibit Gaussian fluctuations. On our scales the nonlinear term still diverges, in fact, it is just a divergent multiple of space–time white noise, as can be seen by considering instead the critically adjusted model

$$\partial_t h_\gamma = \partial_x^2 h_\gamma + \gamma^{1/2}[(\partial_x h_\gamma)^4 - c_{2,\gamma}(\partial_x h_\gamma)^2 + c_{1,\gamma}] + \xi^{(\gamma)}.$$

Although we know of no proof (but see [MU17, GRZ17] for two very recent results going in this direction), taking limits order by order in a perturbative expansion suggests that the limit as  $\gamma \searrow 0$  should just be the free field  $\partial_t h = \partial_x^2 h + a\xi$  with a new  $a > 1$ , suggesting that the solution of (1.5) is essentially the solution of the free equation multiplied by  $\gamma^{-1/2}$ . The lesson is that the only nontrivial limits in (1.4) are going to come from fine tuning  $\varepsilon$  and  $\delta$  with the scale of decay of covariance of the forcing noise. This leads ultimately to two natural choices, the intermediate disorder, and weakly asymmetric limits.

In order to state our results precisely, we need to describe briefly the function spaces we are working in. We would like our initial conditions to have the typical regularity of the KPZ equation, which is  $\mathcal{C}^\alpha$  for  $\alpha < \frac{1}{2}$ , where for  $\alpha \in (0, 1)$  the

Hölder norm is given by

$$\|h\|_\alpha = \|h\|_{L^\infty} + \sup_{x \neq y} \frac{|h(x) - h(y)|}{|x - y|^\alpha}.$$

But even without the noise, it is not at all straightforward to control solutions to (1.4), even for short times, by exploiting the regularization properties of the associated fixed point map. The only tool we really have at our disposal is the maximum principle (see, for example, [BA07]) but it is not clear how one can combine this with the type of analytic estimates essential in the theory of regularity structures.

So we define Hölder spaces  $\mathcal{C}_\varepsilon^{\gamma, \alpha}$  for  $\alpha \in (0, 1)$  and  $\gamma \in (1, 2)$  of functions which are  $\mathcal{C}^\alpha$  at ‘large scales (that is, larger than  $\varepsilon$ ) and  $\mathcal{C}^\gamma$  at ‘small scales’, by setting

$$\|h\|_{\gamma, \alpha; \varepsilon} = \|h\|_\alpha + \sup_{\substack{x \neq y \\ |x - y| \leq \varepsilon}} \frac{|h'(x) - h'(y)|}{\varepsilon^{\alpha - \gamma} |x - y|^{\gamma - 1}}. \quad (1.6)$$

This norm makes such a statement quantitative, typically in the context of a sequence of functions  $h^{(\varepsilon)} \in \mathcal{C}_\varepsilon^{\gamma, \alpha}$  with uniformly bounded norms. For  $\varepsilon = 0$ , one does of course recover the usual  $\alpha$ -Hölder norms. The natural way of comparing an element  $\bar{h} \in \mathcal{C}^\alpha$  with an element  $h \in \mathcal{C}_\varepsilon^{\gamma, \alpha}$  is given by

$$\|h; \bar{h}\|_{\gamma, \alpha; \varepsilon} = \|h - \bar{h}\|_\alpha + \sup_{\substack{x \neq y \\ |x - y| \leq \varepsilon}} \frac{|h'(x) - h'(y)|}{\varepsilon^{\alpha - \gamma} |x - y|^{\gamma - 1}} + \sup_x \frac{|h'(x)|}{\varepsilon^{\alpha - 1}}. \quad (1.7)$$

Note that we do not impose a supremum bound of order  $\varepsilon^{\alpha - 1}$  on  $h'$  in (1.6) because such a bound follows automatically from  $\|h\|_{\gamma, \alpha; \varepsilon} \leq 1$ .

**1.1. Intermediate disorder scaling.** Let us first consider the intermediate disorder regime. In this case,

$$\partial_t h = \partial_x^2 h + F(\partial_x h) + \varepsilon^{1/2} \eta, \quad (1.8)$$

where  $\varepsilon$  will always be a small positive parameter. Setting  $\tilde{h}(x, t) = h(\varepsilon^{-1}x, \varepsilon^{-2}t)$ , we obtain for the rescaled process the equation

$$\partial_t \tilde{h} = \partial_x^2 \tilde{h} + \varepsilon^{-2} F(\varepsilon \partial_x \tilde{h}) + \xi^{(\varepsilon)},$$

where  $\xi^{(\varepsilon)}(x, t) = \varepsilon^{-3/2} \eta(\varepsilon^{-1}x, \varepsilon^{-2}t)$  is a stochastic process that approximates space–time white noise on scales larger than  $\varepsilon$ . Expanding  $F$  in a Taylor series around 0, we formally obtain

$$\partial_t \tilde{h} = \partial_x^2 \tilde{h} + \frac{a_0}{\varepsilon^2} + a_1(\partial_x \tilde{h})^2 + \mathcal{O}(\varepsilon^2(\partial_x \tilde{h})^4) + \xi^{(\varepsilon)}, \quad (1.9)$$

which strongly suggests that the scaling limit of this equation as  $\varepsilon \rightarrow 0$  (modulo a height shift which has the effect of adjusting the value of  $a_0$ ) is given by the KPZ equation [KPZ86].

It also raises the question of what happens if the quadratic part of  $F$  vanishes. Under the scaling given above, it seems intuitively clear that one simply converges towards the ‘trivial’ limit given by the additive stochastic heat equation. On the other hand, one might look at different scalings and consider  $\tilde{h}(x, t) = \varepsilon^\beta h(\varepsilon^{-\alpha}x, \varepsilon^{-2\alpha}t)$  for some exponents  $\alpha$  and  $\beta$  to be determined. Inserting this into (1.8), we obtain the rescaled equation

$$\partial_t \tilde{h} = \partial_x^2 \tilde{h} + \varepsilon^{\beta-2\alpha} F(\varepsilon^{\alpha-\beta} \partial_x \tilde{h}) + \varepsilon^{(1-\alpha+2\beta)/2} \xi^{(\varepsilon^\alpha)}.$$

In order for the noise term to converge to space–time white noise, we should choose  $\beta = (\alpha - 1)/2$ , so that

$$\partial_t \tilde{h} = \partial_x^2 \tilde{h} + \varepsilon^{-(1+3\alpha)/2} F(\varepsilon^{(1+\alpha)/2} \partial_x \tilde{h}) + \xi^{(\varepsilon^\alpha)}. \quad (1.10)$$

If  $F(x) \sim x^{2p}$  around  $x = 0$  for some integer  $p \geq 1$ , this suggests that one should see a nontrivial limit by choosing  $\alpha$  such that  $2p(1 + \alpha) = 1 + 3\alpha$ , that is,  $\alpha = (2p - 1)/(3 - 2p)$  and that the scaling limit should be given by the equation

$$\partial_t \tilde{h} = \partial_x^2 \tilde{h} + (\partial_x \tilde{h})^{2p} + \tilde{\xi},$$

where  $\tilde{\xi}$  denotes space–time white noise. This would to some extent contradict the universality of the KPZ equation. We immediately see a problem with this argument: when  $p > 1$ , the value of  $\alpha$  obtained in this way is negative, so that we do not actually look at large scales at all! We will see that the correct way to rescale this system in order to obtain a nontrivial large-scale limit is to choose  $\alpha = 2p - 1$ . With this choice, it turns out that even if  $p \neq 1$ , the scaling limit obtained in this way is indeed given by the KPZ equation.

In order to fix notations, let us consider henceforth a smooth compactly supported function  $\varrho: \mathbf{R}^2 \rightarrow \mathbf{R}$  integrating to 1 and set

$$\varrho_\varepsilon(t, x) = \varepsilon^{-3} \varrho(\varepsilon^{-2}t, \varepsilon^{-1}x), \quad \xi^{(\varepsilon)} = \varrho_\varepsilon * \xi, \quad (1.11)$$

where ‘ $*$ ’ denotes space–time convolution and  $\xi$  denotes space–time white noise. To keep things simple, we will assume that  $\varrho$  is symmetric in space,  $\varrho(t, x) = \varrho(t, -x)$ . (This is used in a few places such as (6.15) or (6.20). Without the symmetry, one has to make further subtractions, which manifest themselves as global drifts in the resulting equation which then have to be removed by shifts, see [HS15]. In order not to complicate things even further, we do not pursue this here.) Note that, in law, the field  $\xi^{(\varepsilon)}$  is obtained from  $\xi^{(1)}$  as above by a

suitable parabolic rescaling:

$$\xi^{(\varepsilon)}(t, x) \stackrel{\text{law}}{=} \varepsilon^{-3/2} \xi^{(1)}(\varepsilon^{-2}t, \varepsilon^{-1}x).$$

We furthermore define a constant  $C_0$  by

$$C_0 = \iint (P' * Q)(t, x)^2 dt dx, \quad (1.12)$$

where  $P$  denotes the heat kernel on  $[0, 2\pi)$  with periodic boundary conditions. This constant can be rewritten using a graphical notation which will save a great deal of space later. Writing  $\text{-----}$  for the kernel  $Q_\varepsilon * K'$ , a black dot for an integration variable, and a green dot for the value 0, it follows from the definition of  $K$ , the scaling invariance of the heat kernel and the fact that  $Q$  has compact support that one has

$$\text{-----} = \frac{C_0}{\varepsilon} + \mathcal{O}(1). \quad (1.13)$$

We now consider (1.10) with  $\alpha = 2p - 1$ . Performing the substitution  $\varepsilon^{2p-1} \mapsto \varepsilon$ , this can be rewritten as

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + \varepsilon^{-((3p-1)/(2p-1))} F(\varepsilon^{p/(2p-1)} \partial_x h_\varepsilon) + \xi^{(\varepsilon)}. \quad (1.14)$$

As usual, we consider (1.14) on a finite interval with periodic boundary conditions. We now make use of the fact that, by assumption,  $F$  is smooth and  $F(u) \sim u^{2p}$  near  $u = 0$ , so that one can write

$$F(u) = \sum_{k=0}^{2p-1} a_{p+k} u^{2(p+k)} + \tilde{F}(u), \quad (1.15)$$

where  $\tilde{F}$  is a smooth function such that  $|\tilde{F}(u)| \leq |u|^{6p}$  for  $|u| \leq 1$ . Substituting this into (1.14), we obtain the equation

$$\begin{aligned} \partial_t h_\varepsilon &= \partial_x^2 h_\varepsilon + \sum_{k=0}^{2p-1} a_{p+k} \varepsilon^{p-1+((2pk)/(2p-1))} (\partial_x h_\varepsilon)^{2(p+k)} \\ &\quad + \varepsilon^{-((3p-1)/(2p-1))} \tilde{F}(\varepsilon^{p/(2p-1)} \partial_x h_\varepsilon) + \xi^{(\varepsilon)}. \end{aligned} \quad (1.16)$$

With this notation at hand, we have the following result.

**THEOREM 1.1.** Let  $p \geq 1$  be an integer, let  $F \in \mathcal{C}^{6p+1}$  with  $F^{(2p)}(0) \neq 0$  and  $F^{(k)}(0) = 0$  for  $k < 2p$ , and let  $h_\varepsilon$  be the solution to (1.14) with initial condition



$h_0^{(\varepsilon)} \in \mathcal{C}_\varepsilon^{\gamma, \eta}$  for  $\gamma = 2 - \frac{1}{96p}$  and  $\eta \in (\frac{1}{2} - \frac{1}{12p^2}, \frac{1}{2})$  with  $\|h_0^{(\varepsilon)}\|_{\gamma, \eta; \varepsilon} \rightarrow 0$ . Then there exists a constant  $c \in \mathbf{R}$  such that  $h_\varepsilon - (C_\varepsilon + c)t$  converges in probability to  $h_{\text{HC}}^{(\lambda)}$  with initial condition  $h_0$  where

$$\lambda = a_p \frac{C_0^{p-1}(2p)!}{2^p(p-1)!}, \quad C_\varepsilon = \sum_{k=p}^{3p-1} a_k \varepsilon^{-((3p-1-k)/(2p-1))} \frac{C_0^k(2k)!}{2^k k!}. \quad (1.17)$$

**1.2. Weakly asymmetric scaling.** Let us now consider the weakly asymmetric regime

$$\partial_t h = \partial_x^2 h + \sqrt{\varepsilon} F(\partial_x h) + \xi^{(1)}. \quad (1.18)$$

In this case, provided that  $f$  has a nonvanishing second derivative as before, the natural scaling is given by  $\tilde{h}_\varepsilon(x, t) = \varepsilon^{1/2} h(\varepsilon^{-1}x, \varepsilon^{-2}t)$ . With such a scaling, we obtain for  $\tilde{h}$  the equation

$$\partial_t \tilde{h}_\varepsilon = \partial_x^2 \tilde{h}_\varepsilon + \varepsilon^{-1} F(\varepsilon^{1/2} \partial_x \tilde{h}_\varepsilon) + \xi^{(\varepsilon)}. \quad (1.19)$$

Formally replacing  $f$  by its Taylor series as before and neglecting terms of positive order in  $\varepsilon$ , we obtain this time

$$\partial_t \tilde{h}_\varepsilon = \partial_x^2 \tilde{h}_\varepsilon + \varepsilon^{-1} a_0 + a_1 (\partial_x \tilde{h}_\varepsilon)^2 + \mathcal{O}(\varepsilon (\partial_x \tilde{h}_\varepsilon)^4) + \xi^{(\varepsilon)}. \quad (1.20)$$

Comparing this to (1.9), we see that now the ‘error term’ is much larger, so that it is less clear whether this still converges to the KPZ equation. It turns out that it still does, but the ‘error terms’ do not vanish in the limit. Instead, at all orders they contribute to the limiting asymmetry constant  $\lambda$  of the KPZ equation (1.1).

**THEOREM 1.2.** Let  $F: \mathbf{R} \rightarrow \mathbf{R}$  be an even polynomial of degree  $2m$ , let  $\eta \in (\frac{1}{2} - \frac{1}{4m}, \frac{1}{2})$  and  $\gamma = 2 - \frac{1}{32m}$ , and let  $h_0^{(\varepsilon)}$  be a sequence of functions in  $\mathcal{C}_\varepsilon^{\gamma, \eta}$  such that there exists  $h_0 \in \mathcal{C}^\eta$  with  $\lim_{\varepsilon \rightarrow 0} \|h_0^{(\varepsilon)}\|_{\gamma, \eta; \varepsilon} = 0$  in probability. Let  $h_\varepsilon$  be the solution to

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + \varepsilon^{-1} F(\sqrt{\varepsilon} \partial_x h_\varepsilon) + \xi^{(\varepsilon)}, \quad (1.21)$$

where  $\xi^{(\varepsilon)}$  is as in (1.11). Let  $C_0$  be given by (1.12), let  $\mu_0$  be the centred Gaussian measure on  $\mathbf{R}$  with variance  $C_0$ , and let

$$\lambda = \frac{1}{2} \int F''(x) \mu_0(dx), \quad \hat{\lambda} = \int F(x) \mu_0(dx). \quad (1.22)$$

Then there exists a constant  $c$  such that, for every  $T > 0$ , the family of random functions  $(t, x) \mapsto h_\varepsilon(t, x) - (\varepsilon^{-1} \hat{\lambda} + c)t$  converges to  $h_{\text{HC}}^{(\lambda)}$  in probability in  $\mathcal{C}^\eta([0, T] \times S^1)$ .

REMARK 1.3. At this stage, it seems very difficult to obtain uniform moment bounds on solutions to (1.21) as  $\varepsilon \rightarrow 0$ . Therefore, it is unrealistic to expect a much stronger notion of convergence than convergence in probability.

REMARK 1.4. We would like to emphasize again that a naive guess would be that, after being appropriately centred,  $h_\varepsilon$  converges to  $h_{\text{HC}}^{(\lambda)}$  with  $\lambda = a_1$ . It is plain from (1.22) that this is *not* the case. Instead, each of the higher order terms yields a nontrivial contribution in the limit, although they formally disappear as  $\varepsilon \rightarrow 0$  in (1.21). Another remark is that the constant  $\lambda$  which determines the average speed of the interface  $h_\varepsilon$  is in general *different* from  $\varepsilon^{-1}C_0\lambda$ , which is what one would obtain when replacing the nonlinearity by  $\lambda(\partial_x h_\varepsilon)^2$ . Finally, note that the additional constant  $c$  that needs to be subtracted in order to obtain the Hopf–Cole solution depends in a very nontrivial (actually trilinear) way on all of the coefficients of  $P$ .

REMARK 1.5. There are also regimes which seem to combine intermediate disorder and weak asymmetry. For example, if one starts with  $\partial_t h = \partial_x^2 h + \varepsilon^{1/4}\lambda(\partial_x h)^2 + \varepsilon^{1/4}a(\partial_x h)^4 + \varepsilon^{1/4}\xi_\varepsilon$ , then  $\varepsilon^{1/4}h(\varepsilon^{-1}x, \varepsilon^{-2}t) - \tilde{C}_\varepsilon t$  will converge to  $h_{\text{HC}}^{(\lambda)}$  for an appropriate  $\tilde{C}_\varepsilon$ . These regimes are really just versions of intermediate disorder in that in the formal rescaling the higher order error terms are small as in (1.9) as opposed to relevant as in (1.20), and they only contribute to the global shift  $\tilde{C}_\varepsilon$ . Such regimes can be handled using the same methods as intermediate disorder.

REMARK 1.6. A piece of physics lore is that white noise is invariant for the generalized stochastic Burgers equation

$$\partial_t u = \partial_x^2 u + \partial_x F(u) + \partial_x \xi, \quad (1.23)$$

for any polynomial  $F$ . Here one simply thinks of  $u = \partial_x h$  and  $h$  is then a solution to the polynomial KPZ, which, as we learn in this article, simply means quadratic KPZ with a nontrivially renormalized  $\lambda$ . So the invariance of the white noise for (1.23) would appear to be a statement with little new content beyond the white noise invariance for the quadratic case. It is worth remarking however that if we convolve the noise in space only:  $\xi^{(\varepsilon)}(t, x) = \int \xi(t, y)\varrho_\varepsilon(x + y)dy$  where  $\varrho$  is a nonnegative, symmetric function of total integral 1 and  $\varrho_\varepsilon(x) = \varepsilon^{-1}\varrho(\varepsilon^{-1}x)$ , then white noise convolved with  $\varrho_\varepsilon$  is always invariant for the approximating equation

$$\partial_t u_\varepsilon = \partial_x^2 u_\varepsilon + \partial_x C_{2,\varepsilon}(F(u_\varepsilon)) + \partial_x \xi^{(\varepsilon)}, \quad (1.24)$$

where  $C_{2,\varepsilon}f$  denotes convolution with the  $\varepsilon$ -rescaling of  $\varrho_2 = \varrho * \varrho$  (and also the covariance operator of  $\xi^{(\varepsilon)}$ ). This can be shown by adapting [FQ14,

Theorem 2.1], which makes the following argument about the Burgers flow  $\partial_t u_\varepsilon = \partial_x C_{2,\varepsilon}(F(u))$  rigorous:

$$\begin{aligned} \partial_t \int f(u(t)) e^{-(1/2)\langle u, C_{2,\varepsilon}^{-1} u \rangle} &= \int \left\langle \frac{\delta f}{\delta u}, \partial_x C_{2,\varepsilon}(F(u)) \right\rangle e^{-(1/2)\langle u, C_{2,\varepsilon}^{-1} u \rangle} \\ &= - \int f \left\langle \frac{\delta}{\delta u} (\partial_x C_{2,\varepsilon}(F(u)) e^{-(1/2)\langle u, C_{2,\varepsilon}^{-1} u \rangle}) \right\rangle = 0, \end{aligned}$$

where  $\langle f, g \rangle = \int fg dx$  and the last term vanishes because of the following: By Leibniz rule  $(\delta/\delta u)(\partial_x C F e^{-(1/2)\langle u, C^{-1} u \rangle}) = (\delta/\delta u)(\partial_x C F) e^{-(1/2)\langle u, C^{-1} u \rangle} + \partial_x C F (\delta/\delta u)(e^{-(1/2)\langle u, C^{-1} u \rangle})$ . The first term  $(\delta/\delta u) \partial_x C_{2,\varepsilon}(F(u)) = C_{2,\varepsilon}(F''(u) \partial_x u) = \partial_x C_{2,\varepsilon}(F'(u))$  integrates to zero because it is an exact derivative. The second is as well, but this uses the more subtle fact that  $(\delta/\delta u) e^{-(1/2)\langle u, C_{2,\varepsilon}^{-1} u \rangle} = C_{2,\varepsilon}^{-1} u$  and  $\langle \partial_x C_{2,\varepsilon}(F(u)), C_{2,\varepsilon}^{-1} u \rangle = \langle F'(u) \partial_x u, u \rangle = 0$  because if  $G(u) = u F'(u)$  then  $\partial_x G(u) = u F'(u) \partial_x u$ .

**1.3. Possible generalizations.** Although the class of models (1.4) considered in this article is quite rich, we have placed a number of rather severe restrictions on it and it is legitimate to ask whether they are genuinely necessary for our universality result to hold. We now discuss a number of these restrictions and possible strategies for relaxing them.

1. *Regularity of  $F$ .* In the weakly asymmetric limits we assume that  $F$  is a polynomial. The formulation of Theorem 1.2 suggests that this is not an essential assumption since the limiting values of  $\lambda$  and  $\hat{\lambda}$  appearing in the statement are finite for any function (even distribution)  $F$  which is sufficiently tame at infinity. This is a strong hint that it is probably sufficient to impose that  $F$  satisfies a suitable growth condition and is locally Lipschitz continuous. The case of sufficiently smooth  $F$  has now been treated in [HX18], but requires substantial changes in the argument, see also [FG17] for a different approach which should be able to cover the same situation.

The restriction to *even* polynomials is natural because of the lack of a preferred direction, but it is not really important for our proof. Odd polynomials produce large spatial shifts, which simply add a layer of

complication to the argument. It is important to note that we are not using the large-scale convexity of the even polynomial in any way; none of our arguments use convexity at all.

2. *Gaussianity of  $\xi^{(\varepsilon)}$* . At the microscopic level, there is no *a priori* reason for the randomness to be described by Gaussian noise. One may ask whether the arguments in this article still hold if  $\xi$  is an arbitrary smooth and stationary space–time random field with suitable integrability and mixing conditions. (Think of conditions similar to those considered in [PP12, HPP13].) The only part of the paper where we use Gaussianity is in Section 6. In principle, one would expect these results to hold also for suitable non-Gaussian noises (with the same limit). This was done in [HS15] for the particular case when  $F$  is quadratic, but the technique employed there should also work for the general case.
3. *Smoothing mechanism*. One could replace the smoothing mechanism  $\partial_x^2$  in (1.18) by a more general (pseudo-)differential operator of the type  $Q(i\partial_x)$  for an even polynomial (or suitable smooth function)  $Q$ . Provided that  $Q(0) = 0$ ,  $Q''(0) < 0$ ,  $\lim_{|k| \rightarrow \infty} Q(k) = -\infty$ , and  $Q(k) < 0$  for all  $k \neq 0$ , one would expect essentially the same results to still hold true since the large-scale behaviour of the fundamental solution for  $\partial_t - Q(i\partial_x)$  is still described by the heat kernel. Unfortunately, the convergence of the rescaled fundamental solutions does not take place in a topology allowing to easily reuse the results of [Hai14] in this case, although one would still expect the general theory to apply, at least for some choices of  $Q$ .
4. *Symmetry*. Our model is symmetric for the reflection  $x \mapsto -x$ . This symmetry could be broken by considering uneven nonlinearities  $F$  or, in one of the previously discussed generalizations, by considering asymmetric processes  $\xi$  or uneven functions  $Q$ . The expectation is that in this case one should consider limits of the type  $h_\varepsilon(x - c_\varepsilon t, t) - C_\varepsilon t$ , where the constant  $c_\varepsilon$  is also allowed to diverge. The correct choice of these diverging constants should however again lead to the Hopf–Cole solution of the KPZ equation.
5. *The ‘balanced’ weakly asymmetric case*. In the weakly asymmetric case, it may happen that the constant  $\lambda$  in (1.22) is equal to 0. This situation is nongeneric as it requires a very fine balance between all ingredients of the model (since the variance of  $\mu$  in (1.22) depends on the details of both the noise and the smoothing mechanism). In this situation, our results imply that the limiting process is given by the (additive) stochastic heat equation. One might ask whether, similarly to the intermediate disorder case, it is then possible to consider the model on larger scales and still obtain

convergence to KPZ (or some other non-Gaussian process). By analogy with what happens in the context of lattice gases, we do not expect this to be the case [QV13].

6. *Unbounded space.* Our results are on a finite interval with periodic boundary conditions, and extending them to the real line represents a challenge. Recently, [HL15] introduced weighted spaces allowing the extension of the results on convergence of smoothed noise approximations of the quadratic KPZ equation to the whole line. However, these use in an essential way the Hopf–Cole transformation, which is not available for the nonquadratic versions considered in this article.

**1.4. Standing assumptions and terminology.** Throughout the article, we will consider stochastic processes  $h$  taking values in some Banach space  $\mathbf{B}$ . (Typically a space of periodic Hölder continuous functions on  $\mathbf{R}$ .) Since we consider equations with polynomially growing coefficients, we allow for solutions with a finite lifetime  $\tau$  such that  $\lim_{t \rightarrow \tau} \|h(t)\|_{\mathbf{B}} = \infty$  on  $\{\tau < \infty\}$ . One way of formalizing this is to consider, for each  $T > 0$ , the space  $\tilde{\mathcal{C}}_T(\mathbf{B})$  of continuous  $\mathbf{B}$ -valued functions  $h: [0, T] \rightarrow \mathbf{B}$  endowed with a ‘point at infinity’  $\infty$  for which we postulate that

$$d(h, \infty) = d(\infty, h) = \left(1 + \sup_{t \leq T} \|h(t)\|_{\mathbf{B}}\right)^{-1}.$$

For any two elements  $h, \tilde{h} \neq \infty$ , we then set

$$d(h, \tilde{h}) = d(h, \infty) \wedge d(\tilde{h}, \infty) \wedge \sup_{t \leq T} \|h(t) - \tilde{h}(t)\|_{\mathbf{B}}.$$

For fixed  $T$ , we can then view a process  $h$  with lifetime  $\tau$  as a random variable in  $\tilde{\mathcal{C}}_T(\mathbf{B})$  with the understanding that it is equal to  $\infty$  if  $\tau \leq T$ . Throughout the remainder of this article, when we state that a sequence of  $\mathbf{B}$ -valued processes  $h_\varepsilon$  converges in probability to a limit  $h$ , this is a shorthand for the fact that the corresponding  $\tilde{\mathcal{C}}_T(\mathbf{B})$ -valued random variables converge for every choice of final time  $T > 0$ .

Throughout the text, we will make use of the parabolic distance on space–time: if  $z = (t, x)$  then we write  $|z| = \|z\|_s = |t|^{1/2} + |x|$ . We always work on a domain  $z \in [-1, T+1] \times S^1$  where  $S^1 = [0, L)$  with periodic boundary conditions, and we will often write  $\sup_z$  to mean the supremum over  $z$  in this compact set without further comment. The time interval here is chosen to be large enough to strictly contain  $[0, T]$  where convergence results take place.

We will also use  $\lesssim$  throughout to indicate a bound of the left side by a constant multiple of the right side with a constant independent of the relevant quantities. When necessary, these will be indicated explicitly.

## 2. Methodology

In order to prove Theorems 1.1 and 1.2, we make use of the theory of regularity structures as developed in [Hai14, Hai15a]. Let us rapidly recall the main features of this theory. The main idea is to replace the usual  $C^\gamma$  spaces of Hölder continuous functions by analogues  $\mathcal{D}^\gamma$  obtained by extending the usual Taylor polynomials with the addition of a few special universal processes built from the driving noise.

When trying to follow the methodology developed in [Hai14], there are two principal obstacles that must be overcome:

- (1) In (1.19), the parameter  $\varepsilon$  appears in two places: In the regularization of the noise, and multiplying the nonlinearity. If one tries to brutally cast this into the framework of [Hai14], one might try to deal with arbitrary polynomial nonlinearities and view the multiplicative  $\varepsilon$  as simply a parameter of the equation. This is bound to fail since the KPZ equation with a higher than quadratic nonlinearity fails to satisfy the assumption of local subcriticality which is key to the analysis of [Hai14].
- (2) Since polynomials of arbitrary degree are allowed in the right hand side of (1.18), the number of objects that need to be explicitly controlled in the limit  $\varepsilon \rightarrow 0$  can be very large. In the original article [Hai13], almost half of the article was devoted to the control of only five such objects. This was substantially improved in [Hai14], but we heavily exploited the fact that most of the objects that require control for a solution theory to  $\Phi_3^4$  can be decomposed as products of convolutions of integral kernels, for which general bounds exist. In our case, we have to deal with generalized convolutions which cannot be broken into simple convolutions and products.

The second of these is more of a practical nature, and Appendix A contains a very general bound which allows one to control such generalized convolutions, even in the presence of certain renormalization procedures. This bound is then used in Section 6 to give a relatively short proof of the convergence of the required objects as  $\varepsilon \rightarrow 0$ . It has also been used in the article [HP15] to control the necessary objects to provide a Wong–Zakai theorem for a natural class of SPDEs.

The first obstacle above is the main new conceptual difficulty. In a sense, the main point of the regularity structure in [Hai14] is to remove the  $\varepsilon$ -regularization of the noise from the problem: The equation with an arbitrary smooth noise forcing it is simply lifted to the  $\varepsilon$ -independent abstract space. In this way  $\varepsilon$  just takes the role of a parameter in the lifts. However in the present case, the equation itself is also  $\varepsilon$  dependent. So what we want to do is, as much as possible, separate the two  $\varepsilon$ 's.

To accomplish this, we build an extension of the type of regularity structure used in [Hai14, Section 8] which contains an additional abstract symbol  $\mathcal{E}$  representing the multiplicative parameter  $\varepsilon$  appearing in the nonlinearity, but *not* the  $\varepsilon$  in the noise. The resulting regularity structure is described in Section 3 and the corresponding renormalization procedure is described in Section B. To this regularity structure we lift the equation with an arbitrary smooth forcing noise, which does *not* depend on  $\varepsilon$ .

Only in Section 6, when we prove convergence of the models, do we again take the special noise depending on  $\varepsilon$  as in (1.11). For symbols which do not contain  $\mathcal{E}$  this choice is unnecessary. But symbols containing  $\mathcal{E}$  cannot, of course, converge except for this particular choice of approximating noises (or something relatively close).

This turns out to be possible as long as the initial condition is sufficiently smooth. However, the results would then only be valid up to some finite (random) lifetime. To avoid this, we use the fact that the limit can be identified with the Hopf–Cole solution of KPZ, which we know independently is global in time. The difficulty is that one is then forced to start with typical data. This would be slightly below Hölder 1/2, thus leading to a singularity which ruins our fixed point argument. What saves us is that because of the regularization of the noise, the solutions are really smoother on a small scale than this naïve argument suggests. In order to be able to exploit this, we introduce  $\varepsilon$ -dependent versions  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  of the  $\mathcal{D}^{\gamma,\eta}$  spaces, which are generalizations to space–time modelled distributions of weighted Hölder spaces, which, just like the  $\mathcal{C}_\varepsilon^{\gamma,\alpha}$  spaces defined in (1.6), measure regularity differently at scales above and below  $\varepsilon$ . The parameter  $\eta$  appearing here allows for possible blow-up as  $t \rightarrow 0$ , just as in [Hai14, Section 6] (see also Section 4.1). So we cannot completely separate the two  $\varepsilon$ 's, although we try to do it to the largest extent possible.

As we see here, to a certain degree, the  $\varepsilon$  is producing a small scale cutoff in the problem, below which things can be thought of as smooth. This means that multiplication by  $\varepsilon$  effectively increases the homogeneity of a function by 1, and hence our new symbol  $\mathcal{E}$  acts much like an integration operator. There are technical differences however. In the definition of admissible models,  $\Pi_x \mathcal{I} \tau$  is defined in terms of  $\Pi_x \tau$  but  $\Pi_x \mathcal{E} \tau$  is not; in fact, there is much more freedom

in how it is defined. Also,  $\mathcal{E}$  does not need to kill polynomials. More strikingly,  $\mathcal{E}$  is not really even an operator on the regularity structure  $\mathcal{T}$ . The reason is that while we need objects such as  $\mathcal{E}((\mathcal{T}'(\mathcal{E}))^4)$  to describe the right hand side of our equation (where  $\mathcal{E}$  is the lift of the noise), we do *not* need  $(\mathcal{T}'(\mathcal{E}))^4$ , and such an object would not converge, whatever the renormalization.

One unfortunate consequence of these observations is that it makes the structure group highly nontrivial to construct. However, there is a nice trick. We construct a larger regularity structure  $\mathcal{T}_{\text{ex}}$ , which *does* contain objects such as  $(\mathcal{T}'(\mathcal{E}))^4$ , and on which  $\mathcal{E}$  acts much more simply as a linear map defined on a subspace. On this extended regularity structure, the structure group can be constructed as in [Hai14], using the formalism of Hopf algebras. Of course, in  $\mathcal{T}_{\text{ex}}$ , things will not converge in the end, even after renormalization. But our real regularity structure, on which things do converge, is simply a sector of  $\mathcal{T}_{\text{ex}}$ , so the structure group of  $\mathcal{T}_{\text{ex}}$  is defined on it by restriction.

### 3. Construction of the regularity structure

Since the weakly asymmetric case is the more difficult one, we will treat the intermediate disorder scaling essentially as a perturbation of the weakly asymmetric one. The equation of interest is then

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + F_\varepsilon(\partial_x h_\varepsilon) - C_\varepsilon + \xi^{(\varepsilon)}, \quad (3.1)$$

where  $\xi^{(\varepsilon)}$  denotes a regularized version of space–time white noise and the polynomial  $F_\varepsilon$  is of the form

$$F_\varepsilon(u) = \sum_{j=1}^m a_j \varepsilon^{j-1} u^{2j},$$

for some coefficients  $a_j \in \mathbf{R}$  and some finite degree  $m \geq 1$ . Following the methodology of [Hai14], we would like to build a regularity structure that is sufficiently large to be able to accommodate an abstract reformulation of (3.1) as a fixed point problem in some space  $\mathcal{D}$  and which is stable in the limit  $\varepsilon \searrow 0$ .

**3.1. The collection of symbols.** Let us first recall how the construction works for the KPZ equation, where only the term with  $j = 1$  appears in the nonlinearity. In this case, a regularity structure is built in the following way. We write  $\mathcal{U}$  for a collection of symbols, or formal expressions, that will be useful to describe the solution  $h$  as a function of space and time,  $\mathcal{U}'$  for a collection of symbols useful to describe its spatial distributional derivative  $h' = \partial_x h$ , and  $\mathcal{V}$  for a collection of



symbols useful to describe the terms  $F_\varepsilon(\partial_x h_\varepsilon) + \xi^{(\varepsilon)}$  on the right hand side of the KPZ equation. We decree that  $\mathcal{U}$  and  $\mathcal{U}'$  contain at least symbols representing the usual Taylor polynomials, that is, all symbols of the form  $X^k$  for  $k$  a two-dimensional multiindex  $k = (k_1, k_2)$ ,  $k_i \in \{0, 1, 2, \dots\}$ , representing time and space.

Furthermore, we introduce a symbol  $\mathcal{E} \in \mathcal{V}$  describing the driving noise. Finally, we introduce abstract integration maps  $\mathcal{I}$  and  $\mathcal{I}'$  that represent integration with respect to the heat kernel and its spatial derivative, respectively. In view of the structure of the KPZ equation, it is then natural to decree that

$$\begin{aligned} \tau, \bar{\tau} \in \mathcal{U}' &\Rightarrow \tau \bar{\tau} \in \mathcal{V}, \\ \tau \in \mathcal{V} &\Rightarrow \mathcal{I}(\tau) \in \mathcal{U}, \quad \mathcal{I}'(\tau) \in \mathcal{U}', \end{aligned} \quad (3.2)$$

and to define  $\mathcal{U}, \mathcal{U}'$  and  $\mathcal{V}$  as the smallest collection of formal expressions such that  $\mathcal{E} \in \mathcal{V}$ ,  $X^k \in \mathcal{U}$ ,  $X^k \in \mathcal{U}'$ , and (3.2) holds. For consistency with [Hai14], we furthermore decree that  $\mathcal{I}(X^k) = \mathcal{I}'(X^k) = 0$ . In other words, we only keep formal expressions that do not contain  $\mathcal{I}(X^k)$  or  $\mathcal{I}'(X^k)$  as a subexpression. We also decree that  $\tau \bar{\tau} = \bar{\tau} \tau$  and we denote by  $\mathcal{W}$  the union of these collections of formal expressions:

$$\mathcal{W} \stackrel{\text{def}}{=} \mathcal{U} \cup \mathcal{U}' \cup \mathcal{V}.$$

We can then associate to any formal expression  $\tau$  a homogeneity  $|\tau| \in \mathbf{R}$  (despite what the notation may suggest  $|\tau|$  is not necessarily positive) in the following way. For any multiindex  $k = (k_0, k_1)$ , we set  $|X^k| = |k| = 2k_0 + k_1$ . Here,  $k_0$  denotes the degree of the ‘time’ variable, which we choose to count double in order to reflect the parabolic scaling of the heat equation. For the symbol representing the driving noise we set

$$|\mathcal{E}| = -\frac{3}{2} - \kappa, \quad (3.3)$$

where  $\kappa > 0$  is a fixed small value, and we extend this recursively to every formal expression as follows:

$$|\tau \bar{\tau}| = |\tau| + |\bar{\tau}|, \quad |\mathcal{I}(\tau)| = |\tau| + 2, \quad |\mathcal{I}'(\tau)| = |\tau| + 1.$$

With all of these expressions at hand, a simple power-counting argument (see [Hai14, Section 8]) yields the following crucial result.

LEMMA 3.1. *If  $\kappa < \frac{1}{2}$  then for every  $\gamma \in \mathbf{R}$ , the set  $\{\tau \in \mathcal{W} : |\tau| < \gamma\}$  is finite.*

This is a reflection of the fact that the KPZ equation is subcritical with respect to the scaling imposed by the linearized equation. In the context of (3.1),

one could think that it suffices to replace the first implication in (3.2) by

$$\tau_1, \dots, \tau_{2m} \in \mathcal{U}' \Rightarrow \tau_1 \cdots \tau_{2m} \in \mathcal{V}.$$

(Here we exploited the fact that  $\mathbf{1} = X^0$  belongs to  $\mathcal{U}'$ , so that this automatically covers the case of products of less than  $m$  terms.) The problem with this definition is that the conclusion of Lemma 3.1 no longer holds, so that it appears as though the theory developed in [Hai14] breaks down. This is fortunately not the case, but we have to be a little bit more sophisticated.

The reason why we can circumvent the problem is of course that the very singular behaviour of the higher powers of  $\partial_x h$  is precisely compensated by the powers of the small parameter  $\varepsilon$  that multiply them. It is therefore quite reasonable to expect that we can somehow encode this into the properties of our regularity structure. The trick is to introduce an additional symbol  $\mathcal{E}$  besides  $X$ ,  $\mathcal{E}$ ,  $\mathcal{I}$  and  $\mathcal{I}'$  which symbolizes the operation ‘multiplication by  $\varepsilon$ ’. With this new symbol at hand, we build  $\mathcal{U}$ ,  $\mathcal{U}'$  and  $\mathcal{V}$  as before, but we replace the first implication of (3.2) by the implication

$$\tau_1, \dots, \tau_{2k} \in \mathcal{U}' \Rightarrow \mathcal{E}^{k-1} \tau_1 \cdots \tau_{2k} \in \mathcal{V}, \quad (3.4)$$

which we impose for every  $k \in \{1, \dots, m\}$ . The product is made commutative and associative by identifying the corresponding formal expressions and making multiplication by  $\mathbf{1}$  the identity, and  $\mathcal{E}^k \mathcal{E}^\ell \tau = \mathcal{E}^{\ell+k} \tau$ . At this stage, it is very important to note that as a consequence of our definitions, there will be formal expressions  $\tau$  such that  $\mathcal{E}\tau \in \mathcal{W}$ , but  $\tau \notin \mathcal{W}$ . For example,  $\tau = \mathcal{I}'(\mathcal{E})^4$ . This reflects the fact that  $\varepsilon(\partial_x h)^4 - C_\varepsilon$  converges weakly to a distributional limit as  $\varepsilon \rightarrow 0$  for a suitable choice of  $C_\varepsilon$ , while  $(\partial_x h)^4 - C'_\varepsilon$  diverges no matter what is  $C'_\varepsilon$ .

With these notations, we then define  $\mathcal{T}$  as the linear span of  $\mathcal{W}$  and we view the symbols  $\mathcal{E}^{k-1}$  as  $2k$ -linear maps on  $\mathcal{T}$  via

$$(\tau_1, \dots, \tau_{2k}) \mapsto \mathcal{E}^{k-1} \tau_1 \cdots \tau_{2k}.$$

We furthermore decree that the homogeneity of an element of  $\mathcal{W}$  obtained in this way is given by

$$|\mathcal{E}^{k-1} \tau_1 \cdots \tau_{2k}| = k - 1 + \sum_i |\tau_i|. \quad (3.5)$$

Elements  $x \in \mathcal{T}$  can be written uniquely as  $x = \sum_{\tau \in \mathcal{W}} x_\tau \tau$ ,  $x_\tau \in \mathbf{R}$  and with this notation, we set

$$|x|_\alpha = \sum_{|\tau|=\alpha} |x_\tau|, \quad (3.6)$$

with the usual convention that  $|x|_\alpha = 0$  for those  $\alpha$  where the sum is empty.

**3.2. Structure group.** We now describe the structure group  $\mathcal{G}$  associated to the space  $\mathcal{T}$ . For this, we first introduce  $\mathcal{T}_+$ , the free commutative algebra generated by  $\mathcal{W}_+$  which consists of  $X_0, X_1$  as well as the formal expressions

$$\{\mathcal{J}_\ell(\tau) : \tau \in \mathcal{W} \setminus \bar{\mathcal{T}}, |\tau| + 2 > |\ell|\} \cup \{\mathcal{E}_\ell^k(\tau) : \tau \in \mathcal{V}_{\ell,k}\}, \quad (3.7)$$

where  $\ell$  is an arbitrary 2-dimensional multiindex with  $|\ell| = 2\ell_0 + \ell_1$ ,  $k$  is an integer with  $k \in \{1, \dots, m-1\}$ ,  $\bar{\mathcal{T}}$  is the subset generated by the  $X^k$ , and  $\mathcal{V}_{\ell,k}$  is the collection of symbols consisting of  $\tau$  of the form  $\tau_1 \cdots \tau_{2k+2}$ ,  $\tau_i \in \mathcal{U}'$  with  $|\ell| \geq \sum |\tau_i| > |\ell| - k$ . Note that for the moment, elements of  $\mathcal{T}_+$  are formal objects. They are only used to index matrix elements for the linear transformations belonging to the structure group of our regularity structure. The scheme is as follows: Starting from these formal objects, we will define  $\Delta$  by (3.8) and (3.9). The structure group is then defined by (3.10).

**REMARK 3.2.** In principle, there is no *a priori* reason to impose that  $|\ell| \geq \sum |\tau_i|$  in (3.7), through the definition of  $\mathcal{V}_{\ell,k}$  (the analogous constraint does not appear for  $\mathcal{J}_\ell(\tau)$  for example). The reason why have imposed this here is twofold. First, it is natural in view of the canonical lift defined in (3.26) below in the sense that even if we did not impose  $\mathcal{E}_\ell^k(\tau) = 0$  for  $|\ell| < |\tau|$  at the algebraic level, all of the models we ever consider in this article involve linear forms  $f_z$  over  $\mathcal{T}_+$  such that  $f_z(\mathcal{E}_\ell^k(\tau)) = 0$  in that case. The second, more pragmatic, reason is that this greatly simplifies the expression (4.6) which would otherwise sport a number of spurious additional terms.

With this definition at hand, we construct a linear map  $\Delta : \mathcal{T} \rightarrow \mathcal{T} \otimes \mathcal{T}_+$  in a recursive way. In order to streamline notations, we shall write  $\tau^{(1)} \otimes \tau^{(2)}$  as a shorthand for  $\Delta\tau$ . (This is a slight abuse of notation, following Sweedler, since in general  $\Delta\tau$  is a linear combination of such terms. It is justified by the fact that expressions containing the  $\tau^{(i)}$  will always be linear in them.) We then define  $\Delta$  via the identities

$$\Delta \mathbf{1} = \mathbf{1} \otimes \mathbf{1}, \quad \Delta \mathcal{E} = \mathcal{E} \otimes \mathbf{1}, \quad \Delta X_i = X_i \otimes \mathbf{1} + \mathbf{1} \otimes X_i, \quad (3.8)$$

and then recursively by the following relations:

$$\Delta \tau \bar{\tau} = \tau^{(1)} \bar{\tau}^{(1)} \otimes \tau^{(2)} \bar{\tau}^{(2)}, \quad (3.9a)$$

$$\Delta \mathcal{I}(\tau) = \mathcal{I}(\tau^{(1)}) \otimes \tau^{(2)} + \sum_{\ell,k} \frac{X^\ell}{\ell!} \otimes \frac{X^k}{k!} \mathcal{J}_{\ell+k}(\tau), \quad (3.9b)$$

$$\Delta \mathcal{I}'(\tau) = \mathcal{I}'(\tau^{(1)}) \otimes \tau^{(2)} + \sum_{\ell,k} \frac{X^\ell}{\ell!} \otimes \frac{X^k}{k!} \mathcal{J}_{\ell+k+1}(\tau), \quad (3.9c)$$

$$\Delta \mathcal{E}^k(\tau) = \mathcal{E}^k(\tau^{(1)}) \otimes \tau^{(2)} + \sum_{\ell, m} \frac{X^\ell}{\ell!} \otimes \frac{X^m}{m!} \mathcal{E}_{\ell+m}^k(\tau). \quad (3.9d)$$

Here, we write  $\ell + k + 1$  as a shorthand for  $\ell + k + (0, 1)$ , where  $(0, 1)$  is the multiindex corresponding to the spatial direction. We also implicitly set  $\mathcal{J}_k(\tau) = 0$  if  $|\tau| \leq |k| - 2$  and  $\mathcal{E}_\ell^k(\tau) = 0$  if  $|\tau| \leq |\ell| - |k|$  or  $|\tau| > |\ell|$  so these sums, as well as the corresponding ones in the sequel, are all finite.

Finally, we define a linear map  $\mathcal{D}$  on all elements of the type  $X^k \mathcal{I}(\tau)$  by  $\mathcal{D} \mathcal{I}(\tau) = \mathcal{I}'(\tau)$ ,  $\mathcal{D} X^k = k_1 X^{k-(0,1)}$  for every  $k \geq (0, 1)$ ,  $\mathcal{D} \mathbf{1} = 0$ , and by extending it using the Leibnitz rule. It then follows immediately from (3.9b) and (3.9c) that  $\mathcal{D}$  commutes with  $\Delta$  in the sense that  $\Delta \mathcal{D} \tau = (\mathcal{D} \otimes I) \Delta \tau$ .

REMARK 3.3. As already mentioned before (3.5), one should really view the  $\mathcal{E}^{k-1}$  as  $2k$ -multilinear maps. A more pedantic way of writing the last line in the above equation would then be

$$\begin{aligned} \Delta \mathcal{E}^{k-1}(\tau_1, \dots, \tau_{2k}) &= \mathcal{E}^{k-1}(\tau_1^{(1)}, \dots, \tau_{2k}^{(1)}) \otimes \tau_1^{(2)} \cdots \tau_{2k}^{(2)} \\ &\quad + \sum_{\ell, m} \frac{X^\ell}{\ell!} \otimes \frac{X^m}{m!} \mathcal{E}_{\ell+m}^{k-1}(\tau_1, \dots, \tau_{2k}). \end{aligned}$$

However, there is no ambiguity in the above since we implicitly used the fact that  $\Delta$  extends to arbitrary products of elements of  $\mathcal{T}$  via the multiplicative property. This abuse of notation is further justified in view of Section 3.3 below.

For any linear functional  $f: \mathcal{T}_+ \rightarrow \mathbf{R}$ , we can now define in a natural way a map  $\Gamma_f: \mathcal{T} \rightarrow \mathcal{T}$  by

$$\Gamma_f \tau = (I \otimes f) \Delta \tau. \quad (3.10)$$

Let now  $\mathcal{G}_+$  denote the set of all such linear functionals  $f$  which are multiplicative in the sense that  $f(\tau \bar{\tau}) = f(\tau) f(\bar{\tau})$  for any two elements  $\tau, \bar{\tau} \in \mathcal{T}_+$ . With this definition at hand, we set

$$\mathcal{G} = \{\Gamma_f : f \in \mathcal{G}_+\}.$$

It is not difficult to see that these operators are ‘lower triangular’ in the sense that

$$\tau \in \mathcal{T}_\alpha \Rightarrow \Gamma_f \tau - \tau \in \bigoplus_{\beta < \alpha} \mathcal{T}_\beta,$$

but it is not obvious that the set  $\mathcal{G}$  does indeed form a group under composition. In the case where the symbols  $\mathcal{E}^k$  are absent, a proof is given in

[Hai14, Section 8.1]. In our situation, we note that from a purely algebraic point of view, the only thing that distinguishes  $\mathcal{E}^k$  from an abstract integration operator of order  $k$  is that it does not annihilate polynomials. This property was however never used in [Hai14, Section 8.1]. The only reason why this property was imposed in [Hai14] is the aesthetic consideration that we do not want to have a proliferation of abstract symbols that all encode smooth functions, as this would lead to more redundancy in the theory.

REMARK 3.4. While the symbol  $\mathcal{E}$  should be thought as ‘multiplication by  $\varepsilon$ ’ and the models we consider will typically implement this by satisfying the relation (3.26) below, we do not impose that relation in our definition of admissible models. In particular, no real number  $\varepsilon$  needs to be specified in general for the notion of an ‘admissible model’ to make sense. As a matter of fact, while there are natural limiting models with ‘ $\varepsilon = 0$ ’ for which  $\Pi_x \tau = 0$  whenever  $\tau$  contains at least one factor  $\mathcal{E}$ , there are also limiting models for which this is not the case.

REMARK 3.5. We do *not* impose the identity  $\mathcal{I}(\mathcal{E}\tau) = \mathcal{E}\mathcal{I}(\tau)$ , which would in principle have been natural given the interpretation of  $\mathcal{E}$  as essentially multiplication by  $\varepsilon$ . The reason for this is that if we had done this, then we would have run into consistency problems when trying to also impose that  $\mathcal{E}$  increases homogeneity by 1.

**3.3. The extended regularity structure  $\mathcal{T}_{\text{ex}}$ .** Before we proceed, we ‘trim’ the regularity structure  $(\mathcal{T}, \mathcal{G})$  to the bare minimum required for the right hand side of (4.3) to make sense as a map from  $\mathcal{D}'$  into itself for  $\gamma \in (\frac{3}{2} + \kappa, 2 - (6m - 2)\kappa)$ . From now on, with  $\bar{\mathcal{T}}$  the usual Taylor polynomials as before, we set

$$\mathcal{T} = \bar{\mathcal{T}} \oplus \langle \bar{\mathcal{W}} \rangle, \quad \bar{\mathcal{W}} = \bar{\mathcal{U}}' \cup \bar{\mathcal{V}} \cup \{\mathcal{I}(\tau) : \tau \in \bar{\mathcal{V}}\}, \quad (3.11)$$

with

$$\begin{aligned} \bar{\mathcal{U}}' &= \{\tau \in \mathcal{U}' : |\tau| < \tfrac{3}{4}\}, \\ \bar{\mathcal{V}} &= \left\{ \mathcal{E}^{k-1}(\tau_1 \cdots \tau_{2k}) : k \in \{1, \dots, m\}, \tau_i \in \bar{\mathcal{U}}', \sum_{j=1}^{2k} |\tau_j| \leq 0 \right\}, \end{aligned}$$

where we implicitly used the identification  $\mathcal{E}^0(\tau) = \tau$ . Setting furthermore

$$\mathcal{U}'_{\text{ex}} = \{\tau_1 \cdots \tau_{2m} : \tau_i \in \bar{\mathcal{U}}'\}, \quad (3.12)$$

we also define  $\bar{\mathcal{W}}_+$  to consist of  $\{X_0, X_1\}$ , as well as those elements in  $\mathcal{W}_+$  of the form  $\mathcal{J}_\ell(\tau)$  and  $\mathcal{E}_\ell^k(\bar{\tau})$  for elements  $\tau, \bar{\tau} \in \mathcal{W}$  such that  $\mathcal{I}(\tau) \in \bar{\mathcal{W}}$  and  $\bar{\tau} \in \mathcal{U}'_{\text{ex}}$ . With this definition at hand, we define  $\mathcal{T}_+$  as the free commutative algebra generated by  $\bar{\mathcal{W}}_+$ . It will also be very convenient in the sequel to consider an ‘extended’ regularity structure whose structure space  $\mathcal{T}_{\text{ex}}$  is given by

$$\mathcal{T}_{\text{ex}} = \bar{\mathcal{T}} \oplus \langle \mathcal{W}_{\text{ex}} \rangle, \quad \mathcal{W}_{\text{ex}} = \bar{\mathcal{W}} \cup \mathcal{U}'_{\text{ex}}. \quad (3.13)$$

In particular, if we extend the definition of  $\Delta$  to elements in  $\mathcal{U}'_{\text{ex}}$  by imposing that it is multiplicative, our definitions guarantee that

$$\Delta: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}} \otimes \mathcal{T}_+, \quad \Delta: \mathcal{T} \rightarrow \mathcal{T} \otimes \mathcal{T}_+, \quad (3.14)$$

that is, both  $\mathcal{T}$  and  $\mathcal{T}_{\text{ex}}$  are stable under the action of  $\mathcal{G}_+$ , so that  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$  is again a regularity structure and  $(\mathcal{T}, \mathcal{G})$  can be viewed as a sector of  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$ , that is, a subspace that is stable under  $\mathcal{G}$  and diagonal with respect to the direct sum decomposition of  $\mathcal{T}_{\text{ex}}$ . The key point of (3.14) is that the same space  $\mathcal{T}_+$  suffices to define the structure group for  $\mathcal{T}_{\text{ex}}$ , and therefore *the structure group for  $\mathcal{T}_{\text{ex}}$  is the same as the structure group for  $\mathcal{T}$* .

A key advantage of  $\mathcal{T}_{\text{ex}}$ , and this is why we introduce it, is that for  $\ell \leq m-1$ , the maps  $\mathcal{E}^\ell$  can be viewed as genuine *linear* maps defined on the subspace of  $\mathcal{T}_{\text{ex}}$  generated by  $\tau_1 \cdots \tau_{2\ell+2}$ ,  $\tau_i \in \mathcal{U}'$ . However, although  $\mathcal{E}^\ell$  and  $\mathcal{I}$  are defined on subspaces of  $\mathcal{T}_{\text{ex}}$ , they are not necessarily defined for all elements of  $\mathcal{T}_{\text{ex}}$ . The other main advantage of  $\mathcal{T}_{\text{ex}}$  is that all of its elements can be obtained from the ‘basic’ elements  $\{\mathbf{1}, X_0, X_1, \mathcal{E}\}$  by application of the operators  $\mathcal{I}, \mathcal{I}'$  and  $\mathcal{E}^\ell$  without ever leaving  $\mathcal{T}_{\text{ex}}$ . In fact, it is the minimal extension of  $\mathcal{T}$  with that property.

Since  $\mathcal{T} \subset \mathcal{T}_{\text{ex}}$  and since the structure groups are the same for both regularity structures, every model  $(\Pi, \Gamma)$  for  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$  defines a model for  $(\mathcal{T}, \mathcal{G})$  by restriction. (See [Hai14] or Section 3.4.) We will use this fact in Section 3.6 by defining a model first recursively on  $\mathcal{T}_{\text{ex}}$  and then on  $\mathcal{T}$  by restriction. On the other hand, one should remember if  $(\Pi, \Gamma)$  is a model for the structure  $(\mathcal{T}, \mathcal{G})$ , it does *not* automatically extend to a model for  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$ . As a matter of fact, we are precisely interested in the limiting situation in which it does not! Since the structure group  $\mathcal{G}$  is identical for both structures however, the family of operators  $\Gamma_{zz'}$  can be viewed as acting on  $\mathcal{T}_{\text{ex}}$  for any model on  $\mathcal{T}$ . In particular, the spaces  $\mathcal{D}^\gamma$  also make sense over  $\mathcal{T}_{\text{ex}}$  (see Section 3.5 below for the definition of these spaces and their variants and see also Remark 3.14), even for models on  $\mathcal{T}$ . Since  $\mathcal{E}^\ell$  can be thought of as linear operators on  $\mathcal{T}_{\text{ex}}$ , this will give a simple way to understand the fixed point argument.

**3.4. Admissible models.** From now on, we also set  $\mathcal{T} = \bigoplus_{\alpha \in A : \alpha \leq 2} \mathcal{T}_\alpha$ , which has the advantage that  $\mathcal{T}$  is finite-dimensional so we do not need to worry about topologies. In order to describe our ‘polynomial-like’ objects, we first fix a kernel  $K : \mathbf{R}^2 \rightarrow \mathbf{R}$  with the following properties:

- (1) The kernel  $K$  is supported in  $\{|z| \leq 1\}$ ,  $K(t, x) = 0$  for  $t \leq 0$ , and  $K(t, -x) = K(t, x)$ .
- (2) For  $z$  with  $|z| < 1/2$ ,  $K$  coincides with the heat kernel and  $K$  is smooth outside of the origin.
- (3) For every polynomial  $Q : \mathbf{R}^2 \rightarrow \mathbf{R}$  of parabolic degree 2 or higher, one has

$$\int_{\mathbf{R}^2} K(t, x) Q(t, x) dx dt = 0. \quad (3.15)$$

in other words,  $K$  has essentially all the properties of the heat kernel, except that it is furthermore compactly supported and satisfies (3.15). The existence of a kernel  $K$  satisfying these properties is very easy to show.

REMARK 3.6. The identity (3.15) is imposed only for convenience. If we did not impose this, then in order to be able to impose (3.20b) later on we would have to add symbols of the type  $\mathcal{I}(X^k)$  which would also describe smooth functions. This would introduce some rather unnatural redundancy into the construction.

Let  $\mathcal{S}'$  be the space of Schwartz distributions on  $\mathbf{R}^2$  and  $\mathcal{L}(\mathcal{T}, \mathcal{S}')$  the space of linear maps from  $\mathcal{T}$  to  $\mathcal{S}'$ . Furthermore, given a continuous test function  $\varphi : \mathbf{R}^2 \rightarrow \mathbf{R}$  and a point  $z = (t, x) \in \mathbf{R}^2$ , we set

$$\varphi_z^\lambda(\bar{z}) = \lambda^{-3} \varphi((\lambda^{-2}(\bar{t} - t), \lambda^{-1}(\bar{x} - x))),$$

where we also used the shorthand  $\bar{z} = (\bar{t}, \bar{x})$ . Finally, we write  $\mathcal{B}$  for the set of functions  $\varphi : \mathbf{R}^2 \rightarrow \mathbf{R}$  that are smooth, compactly supported in the ball of radius one, and with their values and both first and second derivatives bounded by 1.

Given a kernel  $K$  as above, we then introduce a set  $\mathcal{M}$  of *admissible models* which are analytical objects built upon our regularity structure  $(\mathcal{T}, \mathcal{G})$  that will play a role for our solutions that is similar to that of the usual Taylor polynomials for smooth functions. A *model* (not necessarily admissible) for  $\mathcal{T}$  on  $\mathbf{R}^2$  consists of a pair  $(\Pi, \Gamma)$  of functions

$$\begin{aligned} \Pi : \mathbf{R}^2 &\rightarrow \mathcal{L}(\mathcal{T}, \mathcal{S}') & \Gamma : \mathbf{R}^2 \times \mathbf{R}^2 &\rightarrow \mathcal{G} \\ z &\mapsto \Pi_z & (z, \bar{z}) &\mapsto \Gamma_{z\bar{z}} \end{aligned} \quad (3.16)$$

with the following properties. First, we impose that they satisfy the analytical bounds

$$|(\Pi_z \tau)(\varphi_z^\lambda)| \lesssim \lambda^{|\tau|}, \quad \|\mathcal{Q}_\alpha \Gamma_{z\bar{z}} \tau\| \lesssim |z - \bar{z}|^{|\tau| - \alpha}, \quad (3.17)$$

uniformly over  $\varphi \in \mathcal{B}$ ,  $\lambda \in (0, 1]$ ,  $\tau \in \mathcal{W}$ , and  $\alpha$  with  $\alpha \leq |\tau|$ , where  $\mathcal{Q}_\alpha$  denotes the projection onto  $\mathcal{T}_\alpha$ . Also, the proportionality constants implicit in the notation  $\lesssim$  are assumed to be bounded uniformly for  $z$  and  $\bar{z}$  taking values in any compact set. We furthermore assume that one has the algebraic identities

$$\Pi_z \Gamma_{z\bar{z}} = \Pi_{\bar{z}}, \quad \Gamma_{z\bar{z}} \Gamma_{\bar{z}\bar{z}} = \Gamma_{\bar{z}\bar{z}} \quad (3.18)$$

valid for every  $z, \bar{z}, \bar{\bar{z}}$  in  $\mathbf{R}^2$ .

REMARK 3.7. It is important to note that (3.17) is the crux of the whole theory of regularity structures, providing a concrete meaning to the abstract notion of homogeneity. It is to make (3.17) hold that one is forced to make the subtractions in (3.20c) and (3.20d), which then produces the nontrivial algebraic structure.

In this article, we will always consider *admissible* models that come with some additional structure. Our models will actually consist of pairs  $(\Pi, f)$  where  $\Pi$  is as in (3.16) and  $f: \mathbf{R}^2 \rightarrow \mathcal{G}_+$  is a continuous function such that, if we define

$$\Gamma_{z\bar{z}} = \Gamma_{f_z}^{-1} \Gamma_{f_{\bar{z}}}, \quad (3.19)$$

the properties (3.17) and (3.18) are satisfied. In other words, we assume that there exists one *single* linear map  $\Pi \in \mathcal{L}(\mathcal{T}, \mathcal{S}')$ , where  $\mathcal{S}'$  is the dual of smooth functions, such that  $\Pi_z = \Pi F_z$  for every  $z$ , where  $F_z = \Gamma_{f_z}$ .

DEFINITION 3.8. A model  $(\Pi, f)$  as above is *admissible* on  $\mathcal{T}$  if  $\Pi_z \mathbf{1} = 1$ , for every multiindex  $k$ ,

$$(\Pi_z X^k \tau)(\bar{z}) = (\bar{z} - z)^k (\Pi_z \tau)(\bar{z}), \quad f_z(X^k) = (-z)^k, \quad (3.20a)$$

and, for every  $\tau \in \mathcal{W}$  with  $\mathcal{I}(\tau) \in \mathcal{T}$ , one has the identities

$$f_z(\mathcal{J}_k \tau) = - \int_{\mathbf{R}^2} D^k K(z - \bar{z})(\Pi_z \tau)(d\bar{z}), \quad |k| < |\tau| + 2, \quad (3.20b)$$

$$(\Pi_z \mathcal{I} \tau)(\bar{z}) = \int_{\mathbf{R}^2} K(\bar{z} - \bar{\bar{z}})(\Pi_z \tau)(d\bar{\bar{z}}) + \sum_k \frac{(\bar{z} - z)^k}{k!} f_z(\mathcal{J}_k \tau), \quad (3.20c)$$

$$(\Pi_z \mathcal{I}' \tau)(\bar{z}) = \int_{\mathbf{R}^2} D K(\bar{z} - \bar{\bar{z}})(\Pi_z \tau)(d\bar{\bar{z}}) + \sum_k \frac{(\bar{z} - z)^k}{k!} f_z(\mathcal{J}_{k+1} \tau), \quad (3.20d)$$

where  $D = \partial_x$  and  $k + 1$  means  $(k_0, k_1 + 1)$ .



Note that these definitions in particular also guarantee that  $(\Pi_z \mathcal{D}\tau)(\bar{z}) = \partial_{\bar{z}}(\Pi_z \tau)(\bar{z})$  for every  $\tau$  in the domain of definition of  $\mathcal{D}$ .

REMARK 3.9. Here we set  $\mathcal{J}_k \tau = 0$  if  $|k| \geq |\tau| + 2$ , so that the sum appearing in (3.20c) is always finite. It is not clear in principle that all the integrals appearing in (3.20) converge, but it turns out that the analytical conditions (3.17) combined with the condition  $|k| < |\tau| + 2$  guarantee that this is always the case, see [Hai14, Section 5].

REMARK 3.10. Given an admissible model  $(\Pi, f)$ , we write  $|\Pi|$  for the smallest choice of proportionality constant in (3.17) with the operators  $\Gamma_{z\bar{z}}$  given by (3.19). This is a slight abuse of notation since we should rather write  $|(\Pi, f)|$  instead, but we hope that this notation is lighter while remaining sufficiently unambiguous. Given any two models  $(\Pi, f)$ ,  $(\bar{\Pi}, \bar{f})$ , we furthermore write  $|\Pi; \bar{\Pi}|$  for the same quantity, but with  $\Pi_z$  replaced by  $\Pi_z - \bar{\Pi}_z$  and  $\Gamma_{z\bar{z}}$  replaced by  $\Gamma_{z\bar{z}} - \bar{\Gamma}_{z\bar{z}}$ . Note that these bounds are only locally uniform in general, so these norms also depend on some underlying bounded domain in which we allow  $z$  and  $\bar{z}$  to vary. Since we are only interested in situations with periodic boundary conditions on a bounded domain and on a bounded time interval, this is irrelevant for the purpose of this article. Note also that  $|\cdot|$  is not a norm since the space  $\mathcal{M}$  of admissible models is not linear. It does however behave like a norm for all practical purposes and we will refer to it as the ‘norm’ of a model.

REMARK 3.11. Note that since  $f_z \in \mathcal{G}_+$ , so that it is multiplicative, (3.20a) and (3.20b) do specify  $f_z$  on elements of the type  $\mathcal{J}_k(\tau)$  once we know  $\Pi_z$ . There is therefore quite a lot of rigidity in these definitions, which makes the mere existence of admissible models a highly nontrivial fact.

REMARK 3.12. Building further on Remark 3.11, it actually turns out that if  $\Pi$  satisfies the *first* analytical bound in (3.17) and is such that, for  $F$  defined from  $\Pi$  via (3.20b), one has the identities (3.20c) and (3.20d), then the second analytical bound in (3.17) is *automatically* satisfied for elements of the type  $\mathcal{J}_k(\tau)$  as a consequence of the extension theorem, [Hai14, Theorem 5.14]. However, it is *not* automatic for terms of the type  $\mathcal{E}_\ell^{k-1}(\tau)$ . This is because our notion of an ‘admissible model’ does not specify any relation between  $f_z(\mathcal{E}_\ell^{k-1}(\tau_1, \dots, \tau_{2k}))$  and the distributions  $\Pi_z \tau_i$ .

At this point we have that  $\mathcal{E}$  is an abstract integration operator on the regularity structure  $\mathcal{T}_{\text{ex}}$  and the results of [Hai14, Sections 8.1 and 8.2] hold for  $\mathcal{E}$  by

repeating the proofs there for  $\mathcal{I}$ . These will be used repeatedly in the sequel. Recall however that while  $\mathcal{E}$  is an operator on  $\mathcal{T}_{\text{ex}}$  there are elements of the form  $\mathcal{E}(\tau)$  in  $\mathcal{T}$  for which  $\tau \notin \mathcal{T}$ .

Finally, we define an analogous set  $\mathcal{M}_{\text{ex}}$  of admissible models for  $\mathcal{T}_{\text{ex}}$  on  $\mathbf{R}^2$ . A model for  $\mathcal{T}_{\text{ex}}$  is a pair  $(\Pi, F)$  of functions  $\Pi: \mathbf{R}^2 \rightarrow \mathcal{L}(\mathcal{T}_{\text{ex}}, \mathcal{S}')$  and  $F: \mathbf{R}^2 \rightarrow \mathcal{G}$  satisfying (3.18) and (3.17) for  $\tau \in \mathcal{W}_{\text{ex}}$ , and it is admissible if (3.20a)–(3.20d) hold for  $\tau \in \mathcal{W}$ .

**3.5. Definition of  $\mathcal{D}^\gamma$ .** Consider the space  $\mathcal{T}$  as above and fix some  $\gamma \in \mathbf{R}$ , as well as an admissible model  $(\Pi, F) \in \mathcal{M}$ . We will always assume that  $\gamma$  is different from the (countably many) values that appear as homogeneities in  $\mathcal{T}$ . We then define spaces  $\mathcal{D}^\gamma$  of modelled distributions consisting of those functions  $H: \mathbf{R}^2 \rightarrow \mathcal{T}_{\leq(\gamma \vee 0)} \stackrel{\text{def}}{=} \bigoplus_{\alpha \leq (\gamma \vee 0)} \mathcal{T}_\alpha$  such that

$$\|H\|_\gamma = \sup_{|z-\bar{z}| \leq 1} \sup_{\alpha < \gamma} \frac{|H(z) - \Gamma_{z\bar{z}} H(\bar{z})|_\alpha}{|z - \bar{z}|^{\gamma-\alpha}} + \sup_{\substack{z, \alpha \\ \alpha \leq \gamma \vee 0}} |H(z)|_\alpha < \infty. \quad (3.21)$$

Recall that, as defined in (3.6),  $|H(z)|_\alpha$  refers to the (Euclidean) norm of the part of  $H(z)$  in  $\mathcal{T}_\alpha$ . Here, the arguments  $z, \bar{z}$  are typically constrained to lie furthermore in some fixed bounded set and we have used the shorthand  $\Gamma_{z\bar{z}} \stackrel{\text{def}}{=} F_z^{-1} \circ F_{\bar{z}}$ . Note that the space  $\mathcal{D}^\gamma$  depends on the underlying model! It is however natural to be able to also compare elements  $H$  and  $\bar{H}$  belonging to spaces  $\mathcal{D}^\gamma$  based on two different models  $(\Pi, f)$  and  $(\bar{\Pi}, \bar{f})$ . In this case, we write

$$\begin{aligned} \|H; \bar{H}\|_\gamma &= \sup_{|z-\bar{z}| \leq 1} \sup_{\alpha < \gamma} \frac{|H(z) - \Gamma_{z\bar{z}} H(\bar{z}) - \bar{H}(z) + \bar{\Gamma}_{z\bar{z}} \bar{H}(\bar{z})|_\alpha}{|z - \bar{z}|^{\gamma-\alpha}} \\ &\quad + \sup_{\substack{z, \alpha \\ \alpha \leq \gamma \vee 0}} |H(z) - \bar{H}(z)|_\alpha. \end{aligned} \quad (3.22)$$

This yields a ‘total space’  $\mathcal{M} \ltimes \mathcal{D}^\gamma$  containing all triples of the form  $(\Pi, f, H)$  with  $H \in \mathcal{D}^\gamma$  based on the model  $(\Pi, f)$ . The distances  $|\cdot; \cdot|$  and (3.22) endow  $\mathcal{M} \ltimes \mathcal{D}^\gamma$  with a metric structure.

It was then shown in [Hai14] that for any  $\gamma > 0$  there exists a *unique* locally Lipschitz continuous map  $\mathcal{R}: \mathcal{M} \ltimes \mathcal{D}^\gamma \rightarrow \mathcal{S}'$  with the property that

$$|(\mathcal{R}H - \Pi_z H(z))(\varphi_z^\lambda)| \leq \lambda^\gamma,$$

uniformly over  $\varphi \in \mathcal{B}$ ,  $\lambda \in (0, 1]$  and locally uniformly in  $z$ . The interpretation of the ‘reconstruction operator’  $\mathcal{R}$  is that  $H$  is really just a local description of a ‘Taylor expansion’ for the actual distribution  $\mathcal{R}H$ . It is straightforward to show

that in the particular case where  $\Pi_z \tau$  represents a continuous function for every  $\tau \in T$ , one has the identity

$$(\mathcal{R}f)(z) = (\Pi_z f(z))(z). \quad (3.23)$$

This identity will be crucial in the sequel.

We will also make use of weighted spaces  $\mathcal{D}^{\gamma, \eta}$ , which essentially consist of elements of  $\mathcal{D}^\gamma$  that are allowed to blow up at rate  $\eta$  near the line  $\{(t, x) : t = 0\}$ . For a precise definition, see [Hai14, Definition 6.2]. In our setting, this is the set of functions  $H : \mathbf{R}^2 \rightarrow \mathcal{T}_{\leq (\gamma \vee 0)}$  such that

$$\|H\|_{\gamma, \eta} \stackrel{\text{def}}{=} \sup_{\substack{z, \alpha \\ \alpha \leq \gamma \vee 0}} \frac{|H(z)|_\alpha}{|t|^{((\eta - \alpha) \wedge 0)/2}} + \sup_{|z - \bar{z}| \leq \sqrt{|t| \wedge |\bar{t}|}} \sup_{\alpha < \gamma} \frac{|H(z) - \Gamma_{z\bar{z}} H(\bar{z})|_\alpha}{|z - \bar{z}|^{\gamma - \alpha} (|t| \wedge |\bar{t}|)^{(\eta - \gamma)/2}}, \quad (3.24)$$

where we used  $t$  and  $\bar{t}$  for the time coordinates of  $z$  and  $\bar{z}$ .

REMARK 3.13. Note that we assume  $H(z) \in \mathcal{T}_{\leq (\gamma \vee 0)}$  instead of the more natural assumption  $H(z) \in \mathcal{T}_{< \gamma}$ . This will be useful in the case when  $\gamma < 0$  which we will encounter later on, and where the components in  $\mathcal{T}_\alpha$  for  $\alpha \in [\gamma, 0]$  still play an important role. See in particular Remarks 4.1 and 3.19.

REMARK 3.14. While we have defined the spaces  $\mathcal{D}^\gamma$  and  $\mathcal{D}^{\gamma, \eta}$  for admissible models with respect to  $\mathcal{T}$ , there is of course an analogous definition for admissible models with respect to  $\mathcal{T}_{\text{ex}}$ . Many of the statements in the next several sections will be true for either, and we will indicate if a specific one is being used.

In the limit  $\varepsilon \rightarrow 0$ , we will obtain a model  $(\Pi, F)$  on  $\mathcal{T}$ , not on  $\mathcal{T}_{\text{ex}}$ . However, although  $\Pi$  does not extend to  $\mathcal{T}_{\text{ex}}$ , the operators  $\Gamma_{xy}$  do extend to it by multiplicativity. As a consequence, the spaces  $\mathcal{D}^\gamma$  and  $\mathcal{D}^{\gamma, \eta}$  make sense for function with values in  $\mathcal{T}_{\text{ex}}$ , even if we are only given a model on  $\mathcal{T}$ . If we are given such a model, it is only when applying the reconstruction operator  $\mathcal{R}$  that it is crucial that the function be  $\mathcal{T}$ -valued.

**3.6. Canonical lift to  $\mathcal{T}_{\text{ex}}$ .** Given any *smooth* space–time function  $\zeta$  and any real number  $\varepsilon$ , there is a canonical way of building a family of admissible models,  $\mathcal{L}_\varepsilon(\zeta) = (\Pi^{(\varepsilon)}, f^{(\varepsilon)})$  for the *extended* regularity structure  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$  as follows. First, we set  $\Pi_z^{(\varepsilon)} \mathcal{E} = \zeta$ , independently of  $z$  and of  $\varepsilon$ , and we define it on  $X^k$  as in (3.20a). Then, we define  $\Pi_z^{(\varepsilon)}$  recursively by (3.20c) and (3.20d), together with the identities

$$(\Pi_z^{(\varepsilon)} \tau \bar{\tau})(\bar{z}) = (\Pi_z^{(\varepsilon)} \tau)(\bar{z})(\Pi_z^{(\varepsilon)} \bar{\tau})(\bar{z}) \quad (3.25)$$

as well as

$$(\Pi_z^{(\varepsilon)} \mathcal{E}^{k-1}(\tau))(\bar{z}) = \varepsilon^{k-1} (\Pi_z^{(\varepsilon)} \tau)(\bar{z}) + \sum_{\ell} \frac{(\bar{z} - z)^\ell}{\ell!} f_z^{(\varepsilon)}(\mathcal{E}_\ell^{k-1}(\tau)), \quad (3.26a)$$

$$f_z^{(\varepsilon)}(\mathcal{E}_\ell^{k-1}(\tau)) = -\varepsilon^{k-1} (D^{(\ell)}(\Pi_z^{(\varepsilon)} \tau))(z). \quad (3.26b)$$

None of this makes sense on  $\mathcal{T}$ , which is one of the key reasons to introduce the larger regularity structure  $\mathcal{T}_{\text{ex}}$ . Here, the multiindex  $\ell$  is furthermore constrained by imposing that  $|\tau| \leq |\ell| < k-1+|\tau|$ . Note again that in general, this definition is only guaranteed to make sense if  $\zeta$  is a smooth function! Note also that when we use this definition in practice later on,  $\zeta$  will really be given by some smooth approximation  $\xi_{\bar{\varepsilon}}$  to our space–time white noise. It is however very important to note that  $\bar{\varepsilon}$  can be completely unrelated to  $\varepsilon$ , so the models  $\mathcal{L}_\varepsilon(\xi_{\bar{\varepsilon}})$  or even  $\mathcal{L}_0(\xi_{\bar{\varepsilon}})$  make perfect sense. Finally, note that the definition (3.26) would not even make sense on our actual regularity structure  $\mathcal{T}$ , because we could have  $\mathcal{E}^{k-1}(\tau) \in \mathcal{T}$  but  $\tau \notin \mathcal{T}$ .

PROPOSITION 3.15. If  $\zeta$  is smooth then  $\mathcal{L}_\varepsilon(\zeta) \in \mathcal{M}_{\text{ex}}$  for any  $\varepsilon$ .

*Proof.* The argument is very similar to that of [Hai14, Proposition 8.27]. The fact that the algebraic identity (3.18) is satisfied follows immediately from our construction. The analytical bounds (3.17) for  $\Pi_z$  follow in exactly the same way as in [Hai14, Proposition 8.27] from the stronger bound

$$|(\Pi_z \tau)(\bar{z})| \lesssim |z - \bar{z}|^{|\tau| \wedge 0},$$

which is easily verified by induction. Writing  $\gamma_{z\bar{z}}$  for the element in  $\mathcal{T}_+^*$  such that  $\Gamma_{z\bar{z}} = \Gamma_{\gamma_{z\bar{z}}}$  (recall (3.10) for these notations), the required bounds on  $\Gamma_{z\bar{z}}$  are equivalent to  $\gamma_{z\bar{z}}(\bar{\tau}) \lesssim |z - \bar{z}|^{|\bar{\tau}|}$ . The bounds on  $\gamma_{z\bar{z}}(\bar{\tau})$  for  $\bar{\tau}$  of the form  $\mathcal{I}_k(\tau)$  follow from the bounds on  $\Pi_z$  as in [Hai14, Proposition 8.27], so it only remains to get a bound on  $\gamma_{z\bar{z}}(\mathcal{E}_k^\ell(\tau))$ .

In almost exactly the same way as in the proof of [Hai14, Proposition 8.27], it is straightforward to set up an inductive structure on  $\mathcal{T}$  which allows us to assume that all components of  $\Gamma_{z\bar{z}} \tau$  do satisfy the required bounds. Proceeding exactly as in the last part of the proof of [Hai14, Proposition 8.27], one then obtains the identity

$$\gamma_{z\bar{z}}(\mathcal{E}_k^\ell(\tau)) = f_{\bar{z}}(\mathcal{E}_k^\ell(\tau)) - \sum_m \frac{(\bar{z} - z)^m}{m!} f_z(\mathcal{E}_{k+m}^\ell(\Gamma_{z\bar{z}} \tau)). \quad (3.27)$$

Fix  $\tau$  and write  $g_{\bar{z}}(z)$  as a shorthand for  $-\varepsilon^\ell (D^{(k)}(\Pi_z^{(\varepsilon)} \tau))(z)$ , so that  $f_{\bar{z}}(\mathcal{E}_k^\ell(\tau)) = g_{\bar{z}}(\bar{z})$ . It follows from our construction that the map  $(z, \bar{z}) \mapsto g_{\bar{z}}(z)$

is smooth. It then follows from (3.26) that

$$f_z(\mathcal{G}_{k+m}^\ell(\Gamma_{z\bar{z}}\tau)) = D^{(m)}g_{\bar{z}}(z) \big|_{z=\bar{z}} - \varepsilon^\ell(D^{(k+m)}(\Pi_z^{(\varepsilon)} \text{Proj}_{<|k|+|m|-\ell}\Gamma_{z\bar{z}}\tau))(z), \quad (3.28)$$

where  $\text{Proj}_{<\alpha}$  denotes the orthogonal projection onto  $\mathcal{T}_{<\alpha} \subset \mathcal{T}_{\text{ex}}$ . At this stage, we note that by our induction hypothesis one has  $\|\Gamma_{z\bar{z}}\tau\|_\alpha \lesssim |z - \bar{z}|^{|\tau|-\alpha}$ . In particular, we can combine this with (3.28) to conclude that one has

$$|f_z(\mathcal{G}_{k+m}^\ell(\Gamma_{z\bar{z}}\tau)) - D^{(m)}g_{\bar{z}}(z)| \lesssim |\bar{z} - z|^{|\tau|+\ell-|k+m|}. \quad (3.29)$$

We can also combine (3.28) with (3.27), which yields the identity

$$\gamma_{z\bar{z}}(\mathcal{G}_k^\ell(\tau)) = g_{\bar{z}}(\bar{z}) - \sum_{|m|<|\tau|+\ell-|k|} \frac{(\bar{z} - z)^m}{m!} D^{(m)}g_{\bar{z}}(z), \quad (3.30)$$

which is bounded by a multiple of  $|z - \bar{z}|^{|\tau|+\ell-|k|}$  as a consequence of usual Taylor expansion.  $\square$

It is however very important to keep in mind that not every admissible model is obtained in this way, or even as a limit of such models! This will be apparent in Section B below where we describe the renormalization group.

**3.7. Multiplication by  $\varepsilon^k$ .** For any model that is constructed as the canonical lift of a smooth function as above to  $\mathcal{T}_{\text{ex}}$ , the symbol  $\mathcal{E}^k$  should be thought of as representing the operation of ‘multiplication with  $\varepsilon^k$ ’. This is however not quite true: (3.26a) suggests that we should introduce the (model dependent) linear maps  $\hat{\mathcal{E}}^k$  acting on the spaces  $\mathcal{D}^\gamma$  by

$$(\hat{\mathcal{E}}^k U)(z) = \mathcal{E}^k U(z) - \sum_\ell \frac{X^\ell}{\ell!} f_z(\mathcal{G}_\ell^k(U(z))), \quad (3.31)$$

where  $f$  is determined by the underlying model on which  $\mathcal{D}^\gamma$  is based. One then has the following fact where we implicitly assume that  $U$  takes values in the domain of the operator  $\mathcal{E}^k$ .

**PROPOSITION 3.16.** Let  $\gamma \in \mathbf{R}$  and let  $\delta = \inf\{\gamma - \alpha : \alpha \in A \cap (-\infty, \gamma)\}$ . Then, if  $U \in \mathcal{D}^\gamma$ , one has  $\hat{\mathcal{E}}^k U \in \mathcal{D}^{\bar{\gamma}}$  for  $\bar{\gamma} = (\gamma + k) \wedge \delta$ .

*Proof.* Our aim is to obtain a suitable bound on the components of  $(\hat{\mathcal{E}}^k U)(z) - \Gamma_{zz'}(\hat{\mathcal{E}}^k U)(z')$ . For this, we note that one has from (3.31),

$$\Gamma_{zz'}(\hat{\mathcal{E}}^k U)(z') = \Gamma_{zz'}\mathcal{E}^k U(z') - \sum_\ell \frac{(X + z - z')^\ell}{\ell!} f_{z'}(\mathcal{G}_\ell^k(U(z'))).$$

Now from the analogue for  $\mathcal{E}^k$  of the proof of [Hai14, Theorem 8.24]

$$\Gamma_{zz'} \mathcal{E}^k U(z') = \mathcal{E}^k \Gamma_{zz'} U(z') + \sum_{\ell, m} \frac{X^\ell}{\ell!} \frac{(z - z')^m}{m!} \gamma_{zz'}(\mathcal{E}_{\ell+m}^k(U(z'))). \quad (3.32)$$

At this stage, we make use of the fact that one has the identity [Hai14, p. 127]

$$\gamma_{zz'}(\mathcal{E}_\ell^k \tau) = f_{z'}(\mathcal{E}_\ell^k \tau) - \sum_m \frac{(z' - z)^m}{m!} f_z(\mathcal{E}_{\ell+m}^k \Gamma_{zz'} \tau). \quad (3.33)$$

Inserting this into the above expression and using the binomial identity yields

$$\Gamma_{zz'}(\hat{\mathcal{E}}^k U)(z') = \mathcal{E}^k \Gamma_{zz'} U(z') - \sum_\ell \frac{X^\ell}{\ell!} f_z(\mathcal{E}_\ell^k(\Gamma_{zz'} U(z'))),$$

so that

$$\begin{aligned} (\hat{\mathcal{E}}^k U)(z) - \Gamma_{zz'}(\hat{\mathcal{E}}^k U)(z') &= \mathcal{E}^k(U(z) - \Gamma_{zz'} U(z')) \\ &\quad + \sum_\ell \frac{X^\ell}{\ell!} f_z(\mathcal{E}_\ell^k(\Gamma_{zz'} U(z') - U(z))). \end{aligned} \quad (3.34)$$

The components in  $\mathcal{T}_\alpha$  arising from the first term are bounded by  $|z - z'|^{\gamma+k-\alpha}$  as a trivial consequence of the definition of  $\mathcal{D}^\gamma$  and the fact that  $|\mathcal{E}^k \tau| = |\tau| + k$ , so that we only need to consider the components arising from the second term. For this, we only need to note that these components are bounded by some multiple of  $|z - z'|^\delta$  as an immediate consequence of the definitions of  $\mathcal{D}^\gamma$  and  $\delta$ .  $\square$

REMARK 3.17. There are two very important facts to note here. First, we do *not* assume that  $\gamma > 0$ . Second, the only property of  $f$  that we used is that  $f_z(X^k) = (-z)^k$ . In particular, we do not need to assume that our model is the canonical model associated to a smooth function and parameter  $\varepsilon > 0$ . Actually, we do not even need to assume that it is admissible.

It is then immediate from (3.26) that if this model is the canonical model associated to a smooth function as in Section 3.6, then the reconstruction operator defined in Section 3.5 satisfies the identity

$$\mathcal{R} \hat{\mathcal{E}}^{k-1}(U_1 \cdots U_{2k}) = \varepsilon^{k-1} \mathcal{R} U_1 \cdots \mathcal{R} U_{2k}, \quad (3.35)$$

so that the operation  $\hat{\mathcal{E}}^{k-1}$  does indeed represent multiplication by  $\varepsilon^{k-1}$ .

REMARK 3.18. In general, if we have any model on  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$  consisting of smooth functions and satisfying the identities (3.26), then it is still the case that  $\mathcal{R} \hat{\mathcal{E}}^\ell U = \varepsilon^\ell \mathcal{R} U$ . This remains true even in situations where (3.25) fails and/or

when  $U \in \mathcal{D}^\gamma$  for some  $\gamma < 0$ , provided that in the latter case one considers models consisting of continuous functions and one defines  $\mathcal{R}U$  through the identity  $(\mathcal{R}U)(x) = (\Pi_x U(x))(x)$ .

REMARK 3.19. The operators  $\hat{\mathcal{E}}^k$  will only ever appear in order to describe the right hand side of our equation. By analogy with [Hai14], this suggests that we could in principle restrict ourselves to a regularity structure which only contains those elements of the type  $\mathcal{E}^k \tau$  for which  $|\mathcal{E}^k \tau| < 0$ . Since we furthermore assumed that  $\mathcal{E}_\ell^k(\tau) = 0$  if  $|\tau| \leq |\ell| - |k|$  (see the remark following (3.9d)), this suggests that we could have spared ourselves the trouble of ever introducing the symbols  $\mathcal{E}_\ell^k(\tau)$ .

The reason why this is not the case is that we will apply the operator  $\hat{\mathcal{E}}^k$  to elements  $F \in \mathcal{D}^\gamma$  for some  $\gamma < 0$ , but nevertheless containing nonzero components in spaces  $T_\alpha$  for  $\alpha \in [\gamma, 0]$ , so that  $\hat{\mathcal{E}}^k F$  will have nonzero components for some elements of the type  $\mathcal{E}^k \tau$  with  $|\mathcal{E}^k \tau| > 0$  and for which  $f_z(\mathcal{E}_\ell^k \tau) \neq 0$ . In order to be able to guarantee that  $Q_{\leq 0} \hat{\mathcal{E}}^k F$  still belongs to some space  $\mathcal{D}^\gamma$  with  $\gamma > 0$ , we then still need to control  $\gamma_{z\bar{z}}(\mathcal{E}_\ell^k(\tau))$ .

#### 4. Abstract solution map

We start this section with a computation on  $\mathcal{T}_{\text{ex}}$  showing that if one starts with sufficiently regular initial data, one expects a well-posed fixed point problem in  $\mathcal{D}^\gamma = \mathcal{D}^\gamma(\mathcal{T}_{\text{ex}})$  for  $\gamma > 3/2$ . There are two key issues which will have to be addressed in subsections 4.1–4.4: 1. In order to iterate the argument to get global solutions, we will want to be able to start with less regular initial data; and, 2. We want the fixed point argument on  $\mathcal{T}$  itself, instead of  $\mathcal{T}_{\text{ex}}$  where we can think of  $\mathcal{E}^j$  as abstract integration operators increasing homogeneity by  $j$ . For these reasons we will introduce spaces  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  in Section 4.1. From now on, in order to simplify notations and similarly to [Hai14], we use the shortcut notation

$$\Psi = \mathcal{I}'(\mathcal{E}).$$

We also write  $Q_{\leq 0}$  for the projection onto  $\bigoplus_{\alpha \leq 0} \mathcal{T}_\alpha$  in  $\mathcal{T}_{\text{ex}}$ . Fix now some coefficients  $\hat{a}_j$  and define the linear maps on  $\mathcal{T}_{\text{ex}}$  given by

$$\begin{aligned} \hat{\mathcal{F}}(\tau) &= \sum_{j=1}^m \hat{a}_j Q_{\leq 0} \mathcal{E}^{j-1} (Q_{\leq 0} \Psi^{2j} \tau), \\ \hat{\mathcal{F}}^{(n)}(\tau) &= \sum_{j=[n/2]}^m (2j+1-n) \cdots (2j) \hat{a}_j Q_{\leq 0} \mathcal{E}^{j-1} (Q_{\leq 0} \Psi^{2j-n} \tau), \end{aligned}$$

$$\hat{\mathcal{F}}^{(n)}(\tau) = \sum_{j=\lceil n/2 \rceil}^m (2j+1-n) \cdots (2j) \hat{a}_j \mathcal{E}_0^{j-1} (\mathcal{Q}_{\leq 0} \Psi^{2j-n} \tau).$$

(Of course we assume  $n \leq 2m$ .) We will also write  $\hat{\mathcal{F}}'$  as a shortcut for  $\hat{\mathcal{F}}^{(1)}$  and  $\hat{\mathcal{F}}''$  as a shortcut for  $\hat{\mathcal{F}}^{(2)}$ .

Since the homogeneity of  $\Psi$  is just below  $-1/2$ , and, according to (3.5),  $\mathcal{E}^{j-1}$  increases the homogeneity by  $j-1$ ,  $\hat{\mathcal{F}}$  decreases the homogeneity of its argument by just a bit more than 1,  $\hat{\mathcal{F}}'$  decreases it by just a bit more than  $\frac{1}{2}$ ,  $\hat{\mathcal{F}}''$  decreases it by a little bit more than 0, and all the other  $\hat{\mathcal{F}}^{(n)}$  increase the homogeneity of their argument (provided that  $\kappa$  is small enough). From now on, we will write  $\hat{\mathcal{F}}^{(n)}\tau$  instead of  $\hat{\mathcal{F}}^{(n)}(\tau)$  and we will use the shorthand

$$\hat{\mathcal{F}}^{(n)} = \hat{\mathcal{F}}^{(n)}(\mathbf{1}).$$

Note also that  $\Gamma \hat{\mathcal{F}}^{(n)}(\tau) = \hat{\mathcal{F}}^{(n)}(\Gamma\tau)$  for  $n \leq 2$ , so that one actually has  $\Pi_x \hat{\mathcal{F}} \in \mathcal{C}_s^{-1-2m\kappa}$ ,  $\Pi_x \hat{\mathcal{F}}' \in \mathcal{C}_s^{-(1/2)-(2m-1)\kappa}$ , and so on, for every model  $(\Pi, \Gamma)$ .

Denote now by  $\mathcal{P}$  the integration operator given by

$$\mathcal{P} = \mathcal{K} + R\mathcal{R}, \quad (4.1)$$

where  $\mathcal{K}$  is the operator defined from the kernel  $K$  as in [Hai14, Section 5],  $\mathcal{R}$  is the reconstruction operator, and  $R$  is defined in [Hai14, Lemma 7.7]. For suitable  $\alpha > 0$ , the operator  $\mathcal{P}$  maps  $\mathcal{D}^\alpha$  to  $\mathcal{D}^{\alpha+2}$  as a consequence of [Hai14, Theorem 4.7]. We also write  $\mathbf{1}_+$  for the indicator function of the set of positive times  $\{(t, x) : t > 0\}$ . Because of its discontinuity at the origin, multiplication with  $\mathbf{1}_+$  is not a bounded linear operator on  $\mathcal{D}'$ , so, as in [Hai14], one really does this on  $\mathcal{D}'^{\eta}$  defined at the end of Section 3.5. However, the argument is only formal at this point anyway because of the initial conditions, so we do not pursue it yet. (Note that  $\mathbf{1}_+$  is called  $\mathbf{R}^+$  in [Hai14].)

With these notations at hand it is natural, just as in [Hai14, Section 9], to associate to our problem the fixed point equation

$$H = \mathcal{P}\mathbf{1}_+ \left( \mathcal{E} + \sum_{j=1}^m \hat{a}_j \mathcal{Q}_{\leq 0} \hat{\mathcal{E}}^{j-1} (\mathcal{Q}_{\leq 0} (\mathcal{D}H)^{2j}) \right) + Ph_0, \quad (4.2)$$

where  $\mathcal{D}$  was defined in Section 3.2.

REMARK 4.1. The reason why we are so explicit about the presence of the projection operators  $\mathcal{Q}_{\leq 0}$  (the analogous projections were mostly implicit in [Hai14]) is that we will end up in a situation where  $(\mathcal{D}H)^{2j}$  belongs to a space



$\mathcal{D}^{\gamma_j, \eta_j}$  with  $\gamma_j < 0$  for some  $j$ . Projecting onto  $\mathcal{Q}_{<\gamma_j}$ , as is done in [Hai14], would then have the effect of actually modifying the effect of the reconstruction operator on  $\hat{\mathcal{E}}^{j-1}((\mathcal{D}H)^{2j})$ , which is not a desirable feature.

In principle, one may want to look for solutions to this problem in  $\mathcal{D}^{\gamma, \eta}$  for suitable values of  $\gamma$  and  $\eta$ . The remainder of this section is devoted to the study of (4.2). Before we delve into the details, we give a heuristic argument showing why one would expect this equation to have local solutions. First, we note that (4.2) is of the form

$$H = \mathcal{I}\left(\mathcal{E} + \sum_{j=1}^m \hat{a}_j \mathcal{Q}_{\leq 0} \mathcal{E}^{j-1}(\mathcal{Q}_{\leq 0}(\mathcal{D}H)^{2j})\right) + (\dots), \quad (4.3)$$

where  $(\dots)$  denotes terms taking values in  $\tilde{\mathcal{T}}$ . These additional terms arise as in [Hai14] from the initial condition and from the fact that the operator  $\mathcal{P}$  representing convolution with the heat kernel is given by  $(\mathcal{P}f)(z) = \mathcal{I}f(z) + (\dots)$ , where  $(\dots)$  denotes again some terms taking values in  $\tilde{\mathcal{T}}$ .

It follows that if we are able to solve (4.2) in  $\mathcal{D}^{\gamma, \eta}$  for  $\frac{3}{2} < \gamma < 2 - (6m - 2)\kappa$ , then *any* solution is necessarily of the form

$$H = h \cdot \mathbf{1} + \mathcal{I}(\mathcal{E}) + \mathcal{I}(\hat{\mathcal{F}}) + h' \cdot X + \mathcal{I}(\hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}})) + h' \cdot \mathcal{I}(\hat{\mathcal{F}}'), \quad (4.4)$$

for some continuous real-valued functions  $h = h(t, x)$  and  $h' = h'(t, x)$ . Note that  $h$  is *not* necessarily differentiable and that even when it is,  $h'$  is *not* in general the derivative of  $h$  (see [Hai15b, Section 2] for an introduction and explanation of this issue). This notation is only used by analogy with the usual Taylor expansions. To obtain (4.4), write the right hand side of (4.3) first with  $H = 0$ , then with the resulting  $H$  from the left hand side substituted into the right hand side, and so on, until the expression stabilizes and only components in  $\tilde{\mathcal{T}}$  change from one step to the next. In the simpler context of the KPZ equation, this is explained in the proof of Proposition 15.12 of [FH14]. The abstract derivative of  $H$  is therefore given by

$$\mathcal{D}H = \Psi + \mathcal{I}'(\hat{\mathcal{F}}) + h' \cdot \mathbf{1} + \mathcal{I}'(\hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}})) + h' \mathcal{I}'(\hat{\mathcal{F}}'). \quad (4.5)$$

Regarding the argument of  $\mathcal{I}$  in the right hand side of (4.3), since we only keep terms of negative (or vanishing) homogeneities, it is given by

$$\begin{aligned} & \mathcal{E} + \hat{\mathcal{F}} + \hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}}) + h' \cdot \hat{\mathcal{F}}' + \hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}})) + h' \cdot \hat{\mathcal{F}}' \mathcal{I}'(\hat{\mathcal{F}}') \\ & + \frac{1}{2} \hat{\mathcal{F}}^{(2)}(\mathcal{I}'(\hat{\mathcal{F}}) + h' \cdot \mathbf{1})^2 - \sum_{n \geq 2} f_z(\hat{\mathcal{F}}^{(n)}((\mathcal{I}'(\hat{\mathcal{F}}) + h' \cdot \mathbf{1})^n)) \mathbf{1}. \end{aligned} \quad (4.6)$$

The reason why no other terms of the form  $\hat{\mathcal{F}}^{(n)}(\cdot)$  appear in this expression is that  $\mathcal{E}_0^j(\tau) = 0$  for  $\tau$  such that  $|\tau| > 0$  (see the remark just after (3.9) as well as the definition (4.2) of our fixed point problem).

As a consequence of [Hair14, Theorem 4.7] and Proposition 3.16, we then note that if  $H \in \mathcal{D}^\gamma$  for  $\gamma > \frac{3}{2} + \kappa$  then, disregarding the effect of initial conditions and provided that  $\kappa$  is sufficiently small, the Picard iteration (4.3) maps  $\mathcal{D}^\gamma$  into  $\mathcal{D}^{\gamma'}$  with

$$\gamma' = \gamma + \frac{1}{2} - (2m - 1)\kappa.$$

This strongly suggests that it is possible to build local fixed points of the Picard iteration for  $\kappa$  sufficiently small. It turns out that this heuristic is correct, although technical problems arise due to the effect of the initial condition. The resolution of these problems is the subject of the remainder of this section.

**4.1. Dealing with irregular initial conditions.** There is a problem with the argument outlined above stemming from the class of initial conditions we would like to consider. Since the solutions to the KPZ equation are  $\alpha$ -Hölder continuous only for  $\alpha < \frac{1}{2}$ , we would like to have a (uniform in the small parameter  $\varepsilon$  controlling our smoothing) solution theory for the approximating equations that can deal with this type of initial data. The problem is that in this case, even for fixed  $\varepsilon$ , say  $\varepsilon = 1$ , and considering the deterministic equation

$$\partial_t h = \partial_x^2 h + (\partial_x h)^{2m} + \zeta, \quad h(0, \cdot) = h_0 \in \mathcal{C}^\alpha,$$

for some smooth  $\zeta$ , one expects the supremum norm of  $\partial_x h$  to develop a singularity of order  $t^{(\alpha-1)/2}$  at the origin, since this is what happens for solutions to the heat equation. As a consequence, the term  $(\partial_x h)^{2m}$  leads to a nonintegrable singularity as soon as  $\alpha < 1$  and  $m$  is large enough! ([BA07] gives a nice survey of what is known about the deterministic problem.)

One could of course circumvent this problem by simply postulating that the initial data is smooth (or say Lipschitz continuous). However, in order to obtain approximation results for any fixed time interval, one would like to exploit the global well-posedness of the limiting equation in order to ‘restart’ our approximation argument (see Proposition 4.8). Such an argument would then of course break down since the limiting solutions are only in  $\mathcal{C}^\alpha$  for  $\alpha < \frac{1}{2}$ . On the other hand, it is reasonable to expect the solutions to the approximate equation to remain smooth at scales below  $\varepsilon$ . In order to formalize this, we will introduce spaces of models/functions/modelled distributions that depend on a parameter  $\varepsilon \in (0, 1]$ , as well as their limiting counterparts for  $\varepsilon = 0$ , and we will set up suitable notions of convergence in such a context.

Recall from Section 3 that  $\mathcal{U}' \subset \mathcal{W}$  is the set of all formal expressions in  $\mathcal{W}$  which are of the form  $\mathcal{T}'(\tau)$  for some  $\tau$  in  $\mathcal{W}$ . For  $\varepsilon > 0$ , we then define a class of  $\varepsilon$ -models  $\mathcal{M}_\varepsilon$  which consist of all admissible models  $(\Pi, f)$  that furthermore satisfy the bounds (recall that the sets  $\mathcal{V}_{\ell,k}$  were introduced in Section 3.2)

$$|f_z(\mathcal{E}_\ell^k(\tau))| \leq C\varepsilon^{|\tau|+k-|\ell|}, \quad \tau \in \mathcal{V}_{\ell,k} : |\ell| \geq |\tau| > |\ell| - k, \quad (4.7a)$$

$$|(\Pi_z \tau)(\varphi_z^\lambda)| \leq C\lambda^{\bar{\gamma}} \varepsilon^{|\tau|-\bar{\gamma}}, \quad \tau \in \mathcal{U}', \quad \bar{\gamma} = 1 - \frac{1}{32m}, \quad (4.7b)$$

for some constant  $C$ , uniformly for  $z$  belonging to an arbitrary compact set and for  $\lambda \leq \varepsilon$ . Here,  $m$  is as in (4.2) and  $\kappa$  is as in (3.3). The second bound is assumed to hold uniformly over all test functions  $\varphi \in \mathcal{B}$  as in Section 3.4 such that furthermore  $\int \varphi(z) dz = 0$ .

REMARK 4.2. The second bound in (4.7) is nontrivial (that is, not already implied by the definition of a model) only if  $|\tau| < \bar{\gamma}$ . Note also that the condition on  $\varphi$  guarantees that  $(\Pi_z \mathbf{1})(\varphi_z^\lambda) = 0$ , so that the bound holds trivially for all of  $\tilde{\mathcal{T}}$ .

Note that, viewed as sets, one has of course  $\mathcal{M}_\varepsilon = \mathcal{M}_{\varepsilon'}$  for any  $\varepsilon, \varepsilon' > 0$ . However, they do differ at the level of the corresponding natural distance functions. Indeed, we introduce a natural family of ‘norms’ on  $\mathcal{M}_\varepsilon$  by setting  $|\Pi|_\varepsilon = |\Pi| + \|\Pi\|_\varepsilon$  with

$$\begin{aligned} \|\Pi\|_\varepsilon = & \sup_z \left( \sup_{k,\ell} \sup_{\tau \in \mathcal{V}_{\ell,k}} \varepsilon^{|\ell|-k-|\tau|} |f_z(\mathcal{E}_\ell^k(\tau))| \right. \\ & \left. + \sup_{\substack{\tau \in \mathcal{U}' \\ |\tau| < \bar{\gamma}}} \sup_{\substack{\lambda \leq \varepsilon \\ \varphi}} \lambda^{-\bar{\gamma}} \varepsilon^{\bar{\gamma}+\delta-|\tau|} |(\Pi_z \tau)(\varphi_z^\lambda)| \right), \end{aligned} \quad (4.8)$$

where the supremum over  $\varphi$  runs over the same set as above and  $\delta$  is as in Proposition 3.16. In particular, the restriction of the canonical lift  $\mathcal{L}_\varepsilon(\zeta)$  to  $\mathcal{T}$  is in  $\mathcal{M}_\varepsilon$  for any  $\varepsilon > 0$ .

REMARK 4.3. We have made an abuse of notation here: Unlike for the class of models considered in [Hai14], there is here in general no canonical way of recovering  $f$  from  $\Pi$ , so we should really write  $\|(\Pi, f)\|_\varepsilon$  instead. This is because while our definition of an admissible model imposes (3.20) which determines  $f_z(\mathcal{I}_k \tau)$  in terms of  $\Pi$ , there is no analogue of this for  $f_z(\mathcal{E}_k^\ell \tau)$ . We do have (3.26) for the canonical lift, but this is *not* preserved by our renormalization procedure. Furthermore, unlike (3.20), it is not a continuous relation in the topology on models introduced in [Hai14].

The natural way of comparing two elements of  $\mathcal{M}_\varepsilon$  is to set

$$|\Pi; \bar{\Pi}|_\varepsilon = |\Pi; \bar{\Pi}| + \|\Pi - \bar{\Pi}\|_\varepsilon.$$

The point here is that we will be interested in distance bounds that are uniform in  $\varepsilon$  as  $\varepsilon \rightarrow 0$ .

We also introduce  $\mathcal{M}_0$  which is the subspace of  $\mathcal{M}$  consisting of those admissible models that furthermore satisfy  $f_z(\mathcal{O}_\ell^k(\tau)) = 0$  for every  $\tau$  and every  $k$  and  $\ell$ . Since both  $\mathcal{M}_\varepsilon$  and  $\mathcal{M}_0$  are subspaces of  $\mathcal{M}$ , we can in principle compare them by using the metric  $|\cdot; \cdot|$  on  $\mathcal{M}$ . It will also be convenient to set up a way of comparing elements in  $\mathcal{M}_\varepsilon$  with elements in  $\mathcal{M}_0$  in a way that takes into account the  $\varepsilon$ -dependence. This is done by setting

$$|\Pi; \bar{\Pi}|_{\varepsilon,0} = |\Pi; \bar{\Pi}| + \|\Pi\|_\varepsilon, \quad (4.9)$$

for every pair of admissible models with  $(\Pi, \Gamma) \in \mathcal{M}_\varepsilon$  and  $(\bar{\Pi}, \bar{\Gamma}) \in \mathcal{M}_0$ .

REMARK 4.4. One might wonder if there is a natural way of comparing elements  $(\Pi, \Gamma) \in \mathcal{M}_\varepsilon$  with elements  $(\bar{\Pi}, \bar{\Gamma}) \in \mathcal{M}_{\bar{\varepsilon}}$  for  $0 < \bar{\varepsilon} < \varepsilon$ . For  $\bar{\varepsilon} > \varepsilon/2$  say, it is natural to view both models as belonging to  $\mathcal{M}_\varepsilon$  and to use the distance  $|\cdot; \cdot|_\varepsilon$  defined there. For  $\bar{\varepsilon} < \varepsilon/2$  on the other hand, it is more natural to set  $|\Pi, \bar{\Pi}|_{\varepsilon,\bar{\varepsilon}} = |\Pi; \bar{\Pi}| + \|\Pi\|_\varepsilon + \|\bar{\Pi}\|_{\bar{\varepsilon}}$ . We will however not make use of these definitions in the sequel.

We similarly introduce  $\varepsilon$ -dependent norms on suitable subspaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  of the spaces  $\mathcal{D}^{\gamma,\eta}$  of modelled distributions previously introduced in (3.24). We will usually consider situations where the space  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  is built from an underlying model belonging to  $\mathcal{M}_\varepsilon$ , but this is not needed in general. The space  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  consists of the elements  $H \in \mathcal{D}^{\gamma,\eta}$  such that the norm  $\|H\|_{\gamma,\eta;\varepsilon}$  given by

$$\|H\|_{\gamma,\eta;\varepsilon} = \|H\|_{\gamma,\eta} + \sup_z \sup_{\alpha > \eta} \frac{|H(z)|_\alpha}{\varepsilon^{\eta-\alpha}} + \sup_{\substack{|z-\bar{z}| \leq \sqrt{|r| \wedge |\bar{r}|} \\ |z-\bar{z}| \leq \varepsilon}} \sup_{\alpha < \gamma} \frac{|H(z) - \Gamma_{z\bar{z}} H(\bar{z})|_\alpha}{|z - \bar{z}|^{\gamma-\alpha} \varepsilon^{\eta-\gamma}}, \quad (4.10)$$

is finite.

Note that the space  $\mathcal{D}_0^{\gamma,\eta}$  is nothing but  $\mathcal{D}^{\gamma,\eta}$ . The norms (4.10) are of course all equivalent as long as  $\varepsilon > 0$ , but as  $\varepsilon \rightarrow 0$  they get closer and closer to the inequivalent norm  $\|\cdot\|_{\gamma,\eta}$ .

As before, it is natural to compare elements  $H \in \mathcal{D}_\varepsilon^{\gamma,\eta}$  with elements  $\bar{H} \in \mathcal{D}_0^{\gamma,\eta}$  by setting

$$\begin{aligned} \|H; \bar{H}\|_{\gamma,\eta;\varepsilon} &= \|H; \bar{H}\|_{\gamma,\eta} + \sup_z \sup_{\alpha > \eta} \frac{|H(z)|_\alpha}{\varepsilon^{\eta-\alpha}} \\ &\quad + \sup_{\substack{|z-\bar{z}| \leq \sqrt{|r| \wedge |\bar{r}|} \\ |z-\bar{z}| \leq \varepsilon}} \sup_{\alpha < \gamma} \frac{|H(z) - \Gamma_{z\bar{z}} H(\bar{z})|_\alpha}{|z - \bar{z}|^{\gamma-\alpha} \varepsilon^{\eta-\gamma}}. \end{aligned} \quad (4.11)$$

REMARK 4.5. As before, the fact that  $\bar{H}$  does not appear in the second term of (4.11) is not a typo. Indeed, for general  $\bar{H} \in \mathcal{D}_0^{\gamma,\eta}$  this supremum would in general be infinite.

**4.2. Properties of the spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$ .** In this section, we collect some useful properties of the spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  introduced earlier. Unless otherwise specified, we make the following standing assumptions and abuses of notation:

- Whenever we make a claim of the type ‘if  $H$  belongs to  $\mathcal{D}_\varepsilon^{\gamma,\eta}$ , then  $\bar{H}$  belongs to  $\mathcal{D}_\varepsilon^{\bar{\gamma},\bar{\eta}}$ ’, it is understood that the norm of  $\bar{H}$  can be bounded in terms of the norm of  $H$ , uniformly over  $\varepsilon \in [0, 1]$  and over models in  $\mathcal{M}_\varepsilon$  with bounded norm.
- When comparing modelled distributions in  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  with some in  $\mathcal{D}_0^{\gamma,\eta}$ , we always assume that we are given respective models  $(\Pi, \Gamma) \in \mathcal{M}_\varepsilon$  and  $(\bar{\Pi}, \bar{\Gamma}) \in \mathcal{M}_0$ . Modelled distributions denoted by  $H, H_1$ , and so on, are assumed to belong to spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  based on  $(\Pi, \Gamma)$ , while  $\bar{H}, \bar{H}_1$ , and so on, belong to spaces  $\mathcal{D}_0^{\gamma,\eta}$  based on  $(\bar{\Pi}, \bar{\Gamma})$ .
- Whenever we write  $\Phi \lesssim \Psi$  for two expressions  $\Phi$  and  $\Psi$  depending on  $\varepsilon$ , it is understood that there exists a constant  $C$  independent of  $\varepsilon$  such that  $\Phi \leq C\Psi$ . For every fixed value  $\bar{C} > 0$ , the constant  $C$  can be chosen the same for all possible functions/models appearing in  $\Phi$  and  $\Psi$ , as long as their norms are bounded by  $\bar{C}$ .
- We implicitly assume that the modelled distributions we consider take values in sectors such that the operations we perform are well defined.
- The space–time domain on which our elements are defined is given by  $[0, T] \times S^1$  for some  $T \in [\varepsilon^2, 1]$ .

For all practical purposes, the spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  behave just like the spaces  $\mathcal{D}^{\gamma,\eta}$ . First, we show that the definition (4.10) is somewhat redundant in the sense that the second term is bounded by the two other terms. This shows that in many cases, it suffices to bound the last term in (4.10). Note that this is however *not* the case for (4.11), which is why we chose to keep the current notations.

PROPOSITION 4.6. For  $H \in \mathcal{D}_\varepsilon^{\gamma,\eta}$ , the second term in (4.10) is bounded by a fixed multiple of the sum of the first and the last term.

*Proof.* Since the first term yields  $\|H(t, x)\|_\ell \lesssim t^{(\eta-\ell)/2}$ , the claimed bound is nontrivial only for  $z = (t, x)$  with  $0 < |t| \leq \varepsilon^2$ . For such a value of  $z$ , one can

always find a sequence  $\{z_n\}_{n \geq 0}$  (with  $z_n = (t_n, x_n)$  as usual) such that  $|z_n - z_{n+1}| \leq \sqrt{|t_n| \wedge |t_{n+1}|}$  and  $|z_n - z_{n+1}| \leq \varepsilon c^n$  for some fixed  $c \in (0, 1)$ , and such that  $z_n = z$  for  $n$  sufficiently large. It then suffices to rewrite  $H(z)$  as

$$H(z) = H(z_0) + \sum_{n \geq 0} (H(z_{n+1}) - \Gamma_{z_{n+1}z_n} H(z_n)) + \sum_{n \geq 0} (\Gamma_{z_{n+1}z_n} - 1) H(z_n).$$

The first sum is bounded by a multiple of  $\varepsilon^{\eta-\gamma} \sum_{n \geq 0} |c^n \varepsilon|^{\gamma-\ell}$ , which is the required bound.

To bound the second sum, we proceed by ‘reverse induction’ on  $\ell$ . Indeed, for the largest possible value of  $\ell$  less than  $\gamma$ , one has  $\|(\Gamma_{z_{n+1}z_n} - 1)H(z_n)\|_\ell = 0$ , so that the required bound holds trivially there. Assuming now that the required bound holds for all  $m > \ell$ , we have

$$\|(\Gamma_{z_{n+1}z_n} - 1)H(z_n)\|_\ell \lesssim \sum_{m > \ell} |\varepsilon c^n|^{m-\ell} \|H(z_n)\|_m \lesssim \sum_{m > \ell} |\varepsilon c^n|^{m-\ell} \varepsilon^{\eta-m}.$$

Summing again over  $n$ , the required bound follows.  $\square$

One motivation for our definitions are the following two results. To formulate the first one, we introduce some notation.

**PROPOSITION 4.7.** Let  $\alpha \in (0, 1)$  and  $\gamma \in (1, 2)$ , let  $h \in \mathcal{C}_\varepsilon^{\gamma, \alpha}$ , and let  $Ph$  be the canonical lift (via its truncated Taylor series) of the harmonic extension of  $h$  (in other words, the action of the heat kernel on a function, but then interpreted in the canonical way as a modelled distribution, see [Hai14, (7.13)]). Then, one has  $Ph \in \mathcal{D}_\varepsilon^{\gamma, \alpha}$  and the bound

$$\|Ph\|_{\gamma, \alpha; \varepsilon} \leq C \|h\|_{\gamma, \alpha; \varepsilon}, \quad (4.12)$$

holds uniformly over  $\varepsilon \in [0, 1]$  for some  $C \geq 1$ . If furthermore  $\bar{h} \in \mathcal{C}^\alpha$  and  $h \in \mathcal{C}_\varepsilon^{\gamma, \alpha}$ , then

$$\|Ph; P\bar{h}\|_{\gamma, \alpha; \varepsilon} \leq C \|h; \bar{h}\|_{\gamma, \alpha; \varepsilon}.$$

*Proof.* Since  $\mathcal{C}_\varepsilon^{\gamma, \alpha} \subset \mathcal{C}^\alpha$  with embedding constants uniform in  $\varepsilon$ , we conclude from [Hai14, Lemma 7.5] that we only need to bound the second term in (4.11) (with  $H = Ph$ ). In particular, we only need to consider the case  $\varepsilon > 0$ .

This in turn is nothing but the statement that the map  $Ph$  is of class  $\mathcal{C}^{\gamma/2}$  in time and  $\mathcal{C}^\gamma$  in space, with norm bounded by  $\varepsilon^{\alpha-\gamma}$ . This in turn follows from classical properties of the heat kernel, combined with the fact that the  $\mathcal{C}^\gamma$ -norm of  $h$  is bounded by  $\varepsilon^{\alpha-\gamma}$  by assumption.

To obtain the bound on  $\|Ph; P\tilde{h}\|_{\gamma,\alpha;\varepsilon}$ , we only need to bound the last two terms in (4.11) in terms of the last two terms in (1.7). This follows again immediately from the properties of the heat kernel.  $\square$

We also have the following result, where  $\mathcal{U}$  is as in Section 3.1,  $\langle \mathcal{U} \rangle$  denotes its linear span in  $\mathcal{T}$ , and  $\bar{\gamma}$  is as in (4.7), so that in particular  $\bar{\gamma} > 0$ .

**PROPOSITION 4.8.** Let  $\alpha \leq \frac{1}{2} - \frac{3\kappa}{2}$ , let  $\gamma = 1 + \bar{\gamma}$ , and let  $H_\varepsilon \in \mathcal{D}'$  with values in  $\langle \mathcal{U} \rangle$ , based on some model  $\Pi^{(\varepsilon)} \in \mathcal{M}_\varepsilon$ . Then, for every  $t$  such that  $[t - \varepsilon^2, t + \varepsilon^2] \subset [0, T]$ , the function  $h_t^{(\varepsilon)} = (\mathcal{R}H_\varepsilon)(t, \cdot)$  belongs to  $\mathcal{C}_\varepsilon^{\gamma,\alpha}$  and one has

$$\|h_t^{(\varepsilon)}\|_{\gamma,\alpha;\varepsilon} \leq C \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|_\varepsilon,$$

for some constant  $C$  independent of  $\varepsilon \in (0, 1]$ . Furthermore, given  $H \in \mathcal{D}'$  with values in  $\langle \mathcal{U} \rangle$ , based on some model  $\Pi \in \mathcal{M}_0$ , the function  $h_t = (\mathcal{R}H)(t, \cdot)$  belongs to  $\mathcal{C}^\alpha$  and one has the bound

$$\|h_t; h_t^{(\varepsilon)}\|_{\gamma,\alpha;\varepsilon} \leq C \|H; H_\varepsilon\|_\gamma (|\Pi| + |\Pi^{(\varepsilon)}|_\varepsilon) + |\Pi^{(\varepsilon)}; \Pi|_{\varepsilon,0} (\|H\|_\gamma + \|H_\varepsilon\|_\gamma).$$

*Proof.* Let  $\alpha_0 = |\mathcal{I}(\mathcal{E})| = \frac{1}{2} - \kappa$  be the homogeneity of the element of lowest nonzero homogeneity in  $\mathcal{U}$ . It then follows from [Hai14, Proposition 3.28] that  $\mathcal{R}H_\varepsilon$  is a continuous function with  $\mathcal{R}H_\varepsilon \in \mathcal{C}^{\alpha_0}$  (with parabolic space–time scaling) and, since  $\alpha < \alpha_0$ , that

$$\|h_t^{(\varepsilon)}\|_\alpha \lesssim \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|,$$

so that it only remains to obtain the bound on the last term in (1.6). Setting  $\tilde{h}^{(\varepsilon)}(t, x) = \partial_x h_t^{(\varepsilon)}(x) = (\mathcal{R}\mathcal{D}H_\varepsilon)(t, x)$ , we will prove the stronger fact that  $\tilde{h}^{(\varepsilon)}$  is a continuous function such that

$$\sup_{\substack{z \neq \bar{z} \\ |z - \bar{z}| \leq \varepsilon}} \frac{\varepsilon^{\gamma-\alpha} |\tilde{h}^{(\varepsilon)}(z) - \tilde{h}^{(\varepsilon)}(\bar{z})|}{|z - \bar{z}|^{\gamma-1}} \lesssim \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|_\varepsilon, \quad (4.13)$$

where the supremum runs over  $z$  and  $\bar{z}$  in  $[t - \varepsilon^2/4, t + \varepsilon^2/4] \times S^1$  and  $|z|$  denotes the parabolic distance.

As in [Mey92, Theorem 6.5], the left hand side in (4.13) is bounded, up a factor independent of  $\varepsilon$ , by the quantity

$$\sup_\varphi \sup_{\lambda < \varepsilon} \sup_z \lambda^{1-\gamma} \varepsilon^{\gamma-\alpha} |\tilde{h}^{(\varepsilon)}(\varphi_z^\lambda)|, \quad (4.14)$$

where the first supremum runs over all space–time test functions  $\varphi \in \mathcal{B}$  integrating to 0, the supremum over  $z$  runs over  $[t - \varepsilon^2/2, t + \varepsilon^2/2] \times S^1$ , and  $\tilde{h}^{(\varepsilon)}$  is interpreted as a distribution.

In order to obtain the required bound on (4.14) note that, as a consequence of [Hai14, Lemma 6.7], one has for  $\lambda < \varepsilon$  the bound

$$|(\tilde{h}^{(\varepsilon)} - \Pi_z^{(\varepsilon)} \mathcal{D}H_\varepsilon(z))(\varphi_z^\lambda)| \lesssim \lambda^{\gamma-1} \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}| \leq \varepsilon^{\alpha-\gamma} \lambda^{\gamma-1} \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|_\varepsilon,$$

where we used the fact that  $\gamma > \alpha$  and  $\varepsilon < 1$  to obtain the second inequality. (The reason for requiring the statement of [Hai14, Lemma 6.7] is that we only assume that a temporal neighbourhood of size  $\varepsilon^2$  is included in the domain  $[0, T]$  on which  $H_\varepsilon$  is defined, while the reconstruction theorem [Hai14, Theorem 3.10] would require a neighbourhood of size  $\mathcal{O}(1)$  in order to get bounds uniform in  $\varepsilon$ .) Furthermore, it follows from (4.8), combined with the facts that  $\varphi$  integrates to 0 and  $\bar{\gamma} = \gamma - 1$ , that

$$|(\Pi_z^{(\varepsilon)} \mathcal{D}H_\varepsilon(z))(\varphi_z^\lambda)| \lesssim \lambda^{\gamma-1} \varepsilon^{\alpha_0-\gamma-\delta} \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|_\varepsilon \leq \lambda^{\gamma-1} \varepsilon^{\alpha-\gamma} \|H_\varepsilon\|_\gamma |\Pi^{(\varepsilon)}|_\varepsilon,$$

where  $\delta = \kappa/2$  as above. Here, the second inequality follows from the fact that  $\alpha \leq \alpha_0 - \delta$  by assumption. Combining both of these bounds, the bound (4.14), and therefore also the required bound on  $\|h_t^{(\varepsilon)}\|_{\gamma, \alpha; \varepsilon}$ , follow at once. The bound on  $\|h_t; h_t^{(\varepsilon)}\|_{\gamma, \alpha; \varepsilon}$  then follows in the same way.  $\square$

**4.3. Operations in  $\mathcal{D}_\varepsilon^{\gamma, \eta}$ .** We now show how the basic operations required for our purposes behave in these spaces. First, we have the following bound on the abstract derivatives of modelled distributions:

**PROPOSITION 4.9.** Let  $H \in \mathcal{D}_\varepsilon^{\gamma, \eta}$  for some  $\gamma > 1$  and  $\eta \in \mathbf{R}$ . Then,  $\mathcal{D}H \in \mathcal{D}_\varepsilon^{\gamma-1, \eta-1}$ . Furthermore, one has  $\|\mathcal{D}H; \mathcal{D}\tilde{H}\|_{\gamma-1, \eta-1; \varepsilon} \lesssim \|H; \tilde{H}\|_{\gamma, \eta; \varepsilon}$ .

*Proof.* Immediate from the definitions.  $\square$

We also have a bound on their products:

**PROPOSITION 4.10.** Let  $H_1 \in \mathcal{D}_\varepsilon^{\gamma_1, \eta_1}(V^{(1)})$  and  $H_2 \in \mathcal{D}_\varepsilon^{\gamma_2, \eta_2}(V^{(2)})$  for two sectors  $V^{(1)}$  and  $V^{(2)}$  with respective regularities  $\alpha_1$  and  $\alpha_2$ , such that a product satisfying the properties [Hai14, Definitions 4.1 & 4.6] is defined on  $V^{(1)} \times V^{(2)}$ . Let furthermore  $\gamma = (\gamma_1 + \alpha_2) \wedge (\gamma_2 + \alpha_1)$  and assume that  $\gamma_i > \alpha_i$ . Then, the function  $H = H_1 H_2$  belongs to  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  with  $\eta = (\eta_1 + \alpha_2) \wedge (\eta_2 + \alpha_1) \wedge (\eta_1 + \eta_2)$ .

Furthermore, writing  $H = H_1 H_2$  and  $\tilde{H} = \tilde{H}_1 \tilde{H}_2$ , one has the bound

$$\|H; \tilde{H}\|_{\gamma, \eta; \varepsilon} \lesssim \|H_1; \tilde{H}_1\|_{\gamma_1, \eta_1; \varepsilon} + \|H_2; \tilde{H}_2\|_{\gamma_2, \eta_2; \varepsilon} + |\Pi; \tilde{\Pi}|. \quad (4.15)$$

*Proof.* The proof is identical to that of [Hai14, Proposition 6.12]. The only difference is that when bounding  $H(z) - \Gamma_{z\bar{z}} H(\bar{z})$  one replaces  $\|z; \bar{z}\|_P$  by  $\varepsilon + \sqrt{|t|} \wedge |t'|$  throughout.  $\square$



REMARK 4.11. Note that we did not assume that  $\gamma > 0$ ! In particular, unlike in [Hai14], we do not compose the product with a projection onto  $\mathcal{T}_{<\gamma}$ .

Writing  $Q_{<\alpha}: \mathcal{T} \rightarrow \mathcal{T}$  for the projection onto  $\mathcal{T}_{<\alpha}$ , we also see that such a projection leaves the space  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  invariant.

PROPOSITION 4.12. Let  $F \in \mathcal{D}_\varepsilon^{\gamma,\eta}$  with  $\eta \leq \gamma$  and let  $\alpha \geq \gamma$ . Then, one has again  $Q_{<\alpha}F \in \mathcal{D}_\varepsilon^{\gamma,\eta}$ .

*Proof.* It is sufficient to show that one actually has  $F_\alpha \stackrel{\text{def}}{=} Q_\alpha F \in \mathcal{D}_\varepsilon^{\gamma,\eta}$  for every  $\alpha \geq \gamma$ . It follows from the definitions that  $|F_\alpha(z)| \lesssim (|t| + \varepsilon^2)^{(\eta-\alpha)/2}$ . As a consequence, for  $\beta < \gamma$  (so in particular also  $\beta < \alpha$ ) and for  $|z - \bar{z}| \leq \sqrt{|t| \wedge |\bar{t}|}$ , one has

$$\begin{aligned} |F_\alpha(z) - \Gamma_{z\bar{z}}F_\alpha(\bar{z})|_\beta &\lesssim |z - \bar{z}|^{\alpha-\beta} |F_\alpha(\bar{z})| \lesssim |z - \bar{z}|^{\alpha-\beta} (|t| + \varepsilon^2)^{(\eta-\alpha)/2} \\ &\lesssim |z - \bar{z}|^{\gamma-\beta} (|t| + \varepsilon^2)^{(\eta-\gamma)/2}, \end{aligned}$$

thus yielding the required bound.  $\square$

The following proposition shows how these spaces behave under the action of the integral operator  $\mathcal{K}$  defined in (4.1).

PROPOSITION 4.13. Let  $V$  be a sector of regularity  $\alpha$  and let  $H \in \mathcal{D}_\varepsilon^{\gamma,\eta}(V)$  with  $-2 < \eta < \gamma \wedge \alpha$ . Then, provided that  $\gamma \notin \mathbb{N}$  and  $\eta \notin \mathbb{Z}$ , one has  $\mathcal{K}H \in \mathcal{D}_\varepsilon^{\tilde{\gamma},\tilde{\eta}}$  with  $\tilde{\gamma} = \gamma + 2$  and  $\tilde{\eta} = \eta + 2$ . Furthermore, one has the bound

$$\|\mathcal{K}H; \mathcal{K}\bar{H}\|_{\tilde{\gamma},\tilde{\eta};\varepsilon} \lesssim \|H; \bar{H}\|_{\gamma,\eta;\varepsilon} + |\Pi; \bar{\Pi}|. \quad (4.16)$$

*Proof.* In view of [Hai14, Proposition 6.16] and Proposition 4.6, we only need to bound the last term in (4.10) with  $H$  replaced by  $\mathcal{K}H$ .

This bound follows immediately from the definitions for the components of  $\mathcal{K}H$  that are not proportional to the Taylor monomials, so we only need to consider the latter, that is, we need to show that

$$\|\mathcal{K}H(z) - \Gamma_{z\bar{z}}\mathcal{K}H(\bar{z})\|_\ell \lesssim |z - \bar{z}|^{\tilde{\gamma}-\ell} \varepsilon^{\tilde{\eta}-\tilde{\gamma}},$$

for integer values of  $\ell$  and for  $(z, \bar{z})$  such that  $|z - \bar{z}| \leq \varepsilon \wedge \sqrt{|t| \wedge |\bar{t}|}$ .

The proof of this fact follows the proof of [Hai14, Proposition 6.16] *mutatis mutandis*, so we do not reproduce it here. The only difference is that all the expressions  $\|x, y\|_p$  appearing there are now replaced by  $\varepsilon$ .  $\square$

REMARK 4.14. All conclusions of Proposition 4.13 still hold if  $\mathcal{K}$  is replaced by  $\mathcal{P}$ .

Note that in all the results so far, we never used the fact that the models actually belong to  $\mathcal{M}_\varepsilon$  rather than just  $\mathcal{M}$ . This is somewhat explicit in the fact that the bounds (4.15) and (4.16) depend on  $|\Pi; \bar{\Pi}|$  rather than on  $|\Pi; \bar{\Pi}|_{\varepsilon,0}$ . Furthermore, up to now, while we have seen that the spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  do not behave any ‘worse’ than the spaces  $\mathcal{D}^{\gamma,\eta}$ , they do not behave any ‘better’ either, so it may seem unclear at this stage why we introduced them.

The final property of these spaces that we use is their behaviour under the operation  $\hat{\mathcal{E}}^k$  introduced in Section 3.7. At this stage it is absolutely essential to use the spaces  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  and models in  $\mathcal{M}_\varepsilon$  since the corresponding property would simply be false otherwise.

PROPOSITION 4.15. Let  $H \in \mathcal{D}_\varepsilon^{\gamma,\eta}$  with  $\gamma > -k$  based on a model  $\Pi$  in  $\mathcal{M}_\varepsilon$  and set

$$\bar{\gamma} = \delta, \quad \bar{\eta} = \eta + k,$$

with  $\delta = \inf_{\alpha \in A \cap [-k, \gamma]} (\gamma - \alpha)$ . Then, one has  $\hat{\mathcal{E}}^k H \in \mathcal{D}_\varepsilon^{\bar{\gamma}, \bar{\eta}}$ . Furthermore, for  $\bar{H} \in \mathcal{D}^{\gamma,\eta}$  based on a model  $\bar{\Pi}$  in  $\mathcal{M}_0$ , one has the bound

$$\|\hat{\mathcal{E}}^k H; \hat{\mathcal{E}}^k \bar{H}\|_{\bar{\gamma}, \bar{\eta}; \varepsilon} \lesssim \|H; \bar{H}\|_{\gamma, \eta; \varepsilon} + |\Pi; \bar{\Pi}|_{\varepsilon,0},$$

with a proportionality constant depending on  $|\Pi|_\varepsilon + |\bar{\Pi}|$ , but not explicitly on  $\varepsilon$ .

*Proof.* Setting  $g = \hat{\mathcal{E}}^k H$ , it then follows from (3.34) that

$$g(z) - \Gamma_{zz'} g(z') = \mathcal{E}^k (H(z) - \Gamma_{zz'} H(z')) + f_z(\mathcal{G}_0^k(\Gamma_{zz'} H(z') - H(z))) \mathbf{1}.$$

For the components other than the one multiplying  $\mathbf{1}$ , the required bounds follow at once, provided that  $\bar{\gamma} \leq \gamma + k$  and  $\bar{\eta} \leq \eta + k$ . Regarding the component multiplying  $\mathbf{1}$ , it follows from the definitions of  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  and  $\mathcal{M}_\varepsilon$  that the terms arising from components of  $\Gamma_{zz'} H(z') - H(z)$  proportional to  $\tau$  are bounded by

$$\varepsilon^{|\tau|+k} |z - z'|^{\gamma-|\tau|} (\varepsilon + \sqrt{|t|})^{\eta-\gamma}, \quad (4.17)$$

where  $t$  is the time component of  $z$  and we only consider pairs  $z, z'$  such that  $|z - z'|^2 \leq |t|/2$ , say. If  $|z - z'| \leq \varepsilon$ , then this bound gets worse for larger values of  $|\tau|$ . By the definition of  $\delta$  the largest value that arises is given by at most  $|\tau| = \gamma - \delta$ . It follows that the requested bound holds, provided that  $\bar{\gamma} \leq \delta$  and

$\bar{\eta} \leq \delta + \eta - \gamma$ . For  $|z - z'| \geq \varepsilon$ , the bound (4.17) is worse for small values of  $\tau$ . Since the smallest possible value of  $\tau$  contributing to it is  $|\tau| = \delta - k$ , this expression is bounded by  $|z - z'|^{\gamma+k} (\varepsilon + \sqrt{|t|})^{\eta-\gamma}$ . Since furthermore we only consider pairs  $z, z'$  such that  $|z - z'| \leq \varepsilon + \sqrt{|t|}$ , this is also bounded by a multiple of  $|z - z'|^\delta (\varepsilon + \sqrt{|t|})^{\eta+k-\delta}$  as required.

We now turn to the pointwise bound on  $g$ . For the components not multiplying **1**, it is immediate to see that the required bound holds as soon as  $\bar{\eta} \leq \eta + k$ . The component multiplying **1** is given by  $f_z(\mathcal{E}_0^k(H(z)))$ . Again, the worst available bound is on the component of  $H(z)$  multiplying  $\tau$  with  $|\tau| = \gamma - \delta$ , for which we obtain a bound of the type

$$|\langle g(z), \mathbf{1} \rangle| \lesssim |f_z(\mathcal{E}_0^k(\tau))| (\varepsilon + \sqrt{|t|})^{(\eta-|\tau|) \wedge 0} \|H\|_{\gamma, \eta; \varepsilon}.$$

At this stage, we make use of the assumption that the underlying model belongs to  $\mathcal{M}_\varepsilon$ , which guarantees that

$$|f_z(\mathcal{E}_0^k(\tau))| \lesssim \varepsilon^{|\tau|+k}. \quad (4.18)$$

Since only terms with  $|\tau| + k > 0$  contribute (see the remark following (3.9)), we conclude that

$$|\langle g(z), \mathbf{1} \rangle| \lesssim (\varepsilon + \sqrt{|t|})^{(\eta+k) \wedge 0} \|H\|_{\gamma, \eta; \varepsilon} \leq (\varepsilon + \sqrt{|t|})^{\bar{\eta} \wedge 0} \|H\|_{\gamma, \eta; \varepsilon}, \quad (4.19)$$

provided that  $\bar{\eta} < \eta + k$ , which is the required bound.

It remains to bound  $\|\hat{\mathcal{E}}^k H; \hat{\mathcal{E}}^k \bar{H}\|_{\bar{\gamma}, \bar{\eta}; \varepsilon}$ . For this, the bounds on  $\|\hat{\mathcal{E}}^k H; \hat{\mathcal{E}}^k \bar{H}\|_{\bar{\gamma}, \bar{\eta}}$  follow in the same way as above. The bound on the second term in (4.11) also follows in the same way, noting that it only requires the bounds (4.18) which in turn are controlled by  $|\Pi; \bar{\Pi}|_{\varepsilon, 0}$  as a consequence of (4.9) and (4.8).  $\square$

**4.4. Picard iteration and convergence.** We now show that the ‘abstract’ fixed point problem associated to our equation is uniformly well behaved in the spaces  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  for suitable values of  $\gamma$  and  $\eta$ . (This is precisely what motivates our choice of definitions for  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  in the first place.) More precisely, we have the following result.

**THEOREM 4.16.** Let  $m \geq 1$ ,  $\eta \in (\frac{1}{2} - \frac{1}{4m}, \frac{1}{2})$ ,  $\varepsilon \in [0, \varepsilon_0]$ , and let  $\kappa > 0$  be sufficiently small (depending only on  $m$  and  $\eta$ ). Let furthermore  $\gamma = 2 - \nu$  with  $\nu = 1/(32m)$ , and consider the fixed point equation

$$H = \mathcal{P}\mathbf{1}_+ \left( \mathcal{E} + \sum_{j=1}^m \hat{a}_j \mathcal{Q}_{\leq 0} \hat{\mathcal{E}}^{j-1} (\mathcal{Q}_{\leq 0} (\mathcal{D}H)^{2j}) \right) + Ph_0, \quad (4.20)$$

for some  $h_0 \in \mathcal{C}_\varepsilon^{\gamma,\eta}$ . Then, for  $\varepsilon \leq \varepsilon_0$  with  $\varepsilon_0$  and the final time  $T > 0$  sufficiently small and for any model in  $\mathcal{M}_\varepsilon$ , there exists a unique solution to (4.20) in  $\mathcal{D}_\varepsilon^{\gamma,\eta}$ . Furthermore, the time  $T$  can be chosen uniformly over bounded sets of initial conditions in  $\mathcal{C}_\varepsilon^{\gamma,\eta}$ , over bounded sets in  $\mathcal{M}_\varepsilon$ , over bounded sets in the space of parameters  $\hat{a}_1, \dots, \hat{a}_m$ , and over  $\varepsilon \in [0, \varepsilon_0]$ .

Let  $h_0^{(\varepsilon)} \in \mathcal{C}_\varepsilon^{\gamma,\eta}$  be a sequence of elements such that there exists  $h_0 \in \mathcal{C}^\eta$  with  $\lim_{\varepsilon \rightarrow 0} \|h_0^{(\varepsilon)}; h_0^{(\varepsilon)}\|_{\gamma,\eta;\varepsilon} = 0$ , and let  $\Pi^{(\varepsilon)} \in \mathcal{M}_\varepsilon$  be a sequence of models such that there exists  $\Pi \in \mathcal{M}_0$  with  $\lim_{\varepsilon \rightarrow 0} |\Pi^{(\varepsilon)}; \Pi|_{\varepsilon,0} = 0$ . Let  $T > 0$  be fixed and assume that  $H \in \mathcal{D}_0^{\gamma,\eta}$  solves (4.20) with model  $\Pi$  up to some terminal time  $T > 0$ . Then, for  $\varepsilon > 0$  small enough, there exists a unique solution  $H_\varepsilon \in \mathcal{D}_\varepsilon^{\gamma,\eta}$  to (4.20) with initial condition  $h_0^{(\varepsilon)}$  and model  $\Pi_\varepsilon$  up to time  $T$ , and  $\lim_{\varepsilon \rightarrow 0} \|H^{(\varepsilon)}; H\|_{\gamma,\eta;\varepsilon} = 0$ .

*Proof.* We first prove that the fixed point problem (4.20) can be solved locally with dependencies of the local existence time that are uniform in  $\varepsilon$ , provided that both the initial condition and the underlying model are controlled in the corresponding  $\varepsilon$ -dependent norms. We consider (4.20) as a fixed point argument in  $\mathcal{D}_\varepsilon^{\gamma,\eta}$ . In other words, we show that if we denote by  $\mathcal{M}$  the map

$$\mathcal{M}(H) = \mathcal{P}\mathbf{1}_+ \left( \mathcal{E} + \sum_{j=1}^m \hat{a}_j \mathcal{Q}_{\leq 0} \hat{\mathcal{E}}^{j-1} (\mathcal{Q}_{\leq 0} (\mathcal{D}H)^{2j}) \right) + Ph_0^{(\varepsilon)}, \quad (4.21)$$

then, for sufficiently small values of the final time  $T$  and uniformly in the stated data,  $\mathcal{M}$  is a contraction mapping the centred ball of large enough radius  $R$  in  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  into the ball of radius  $R/2$ . Additional details, in particular the proof that solutions can be continued uniquely until the explosion time in  $\mathcal{C}_\varepsilon^\eta$ , can be found in [Hai14, Section 7].

Regarding the term  $Ph_0^{(\varepsilon)}$ , it follows from Proposition 4.7, combined with our assumptions on the initial conditions, that it belongs to  $\mathcal{D}_\varepsilon^{\gamma,\eta}$ , uniformly over  $\varepsilon \in [0, 1]$ , and that  $\|Ph_0^{(\varepsilon)}; Ph_0\|_{\gamma,\eta;\varepsilon} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ .

Combining Propositions 4.9, 4.10 and 4.15, we conclude that if we set

$$\gamma_1 = \gamma - \frac{1}{2} - j - \kappa(2j - 1), \quad \eta_1 = 2j(\eta - 1),$$

then the map  $H \mapsto (\mathcal{D}H)^{2j}$  is continuous from  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  into  $\mathcal{D}_\varepsilon^{\gamma_1,\eta_1}$ . Note that  $\gamma_1$  is negative as soon as  $j \geq 2$ , so that by Proposition 4.12 the map  $H \mapsto \mathcal{Q}_{\leq 0} (\mathcal{D}H)^{2j}$  is also continuous from  $\mathcal{D}_\varepsilon^{\gamma,\eta}$  into  $\mathcal{D}_\varepsilon^{\gamma_1,\eta_1}$  as soon as  $j \geq 2$ . For  $j = 1$ , it turns out that one actually has  $\mathcal{Q}_{\leq 0} (\mathcal{D}H)^2 = \mathcal{Q}_{< \gamma_1} (\mathcal{D}H)^2$  as a consequence of the fact that  $\gamma_1 < \frac{1}{2} - \frac{1}{32m}$  and the homogeneities appearing in  $\mathcal{T}_{\text{ex}}$  are arbitrarily close (from below) to half-integers when  $\kappa$  is small, so that this term also belongs to  $\mathcal{D}_\varepsilon^{\gamma_1,\eta_1}$ .

Since the homogeneities of elements of  $\mathcal{W}$  with homogeneity smaller than 2 (say) are all of the form  $\frac{k}{2} - \ell\kappa$  for  $k$  and  $\ell$  some integers with  $\ell$  bounded by some fixed multiple of  $m$ , we can apply Proposition 4.15 with  $\delta = \frac{1}{2} - 2\nu$  provided that we choose  $\kappa$  sufficiently small. As a consequence, we see that  $H \mapsto \hat{\mathcal{E}}^{j-1}(Q_{\leq 0}(\mathcal{D}H)^{2j})$  is continuous from  $\mathcal{D}_{\varepsilon}^{\gamma, \eta}$  into  $\mathcal{D}_{\varepsilon}^{\delta, \eta_2}$  with

$$\eta_2 = j(2\eta - 1) + \kappa(2j - 1) + \frac{1}{2} + \delta - \gamma = j(2\eta - 1) + \kappa(2j - 1) - 1 - \nu.$$

In order to be able to apply Proposition 4.13, we would like to guarantee that  $\eta_2 > -2$ . Provided that  $\kappa$  is sufficiently small, this is the case if  $j(2\eta - 1) > -1 + 2\nu$  for  $j \leq m$  which, keeping in mind our choice of  $\nu$ , is guaranteed by the condition  $\eta > \frac{1}{2} - \frac{1}{4m}$ .

It then follows from Propositions 4.13 and 4.12 that, again provided that  $\kappa$  is chosen sufficiently small, there exists  $\theta > 0$  such that  $\mathcal{P}\mathbf{1}_+ Q_{\leq 0} \hat{\mathcal{E}}^{j-1}(Q_{\leq 0}(\mathcal{D}H)^{2j})$  belongs to  $\mathcal{D}_{\eta}^{\gamma, \eta+\theta}$ , provided that

$$j(2\eta - 1) + \kappa(2j - 1) + 1 - \nu \geq \eta + \theta.$$

This is the case if  $\eta(2j - 1) > j - 1 + 2\nu$  for  $j = 1, \dots, m$ , which in turn is again guaranteed by the assumption that  $\eta > \frac{1}{2} - \frac{1}{4m}$ . Since the heat kernel is nonanticipative, we actually know a little bit more: as a consequence of [Hai14, Theorem 7.1, Lemma 7.3], we know that

$$\|\mathcal{P}\mathbf{1}_+ H\|_{\gamma, \eta} \leq CT^{\theta} \|H\|_{\delta, \eta_2; \varepsilon},$$

where  $T$  denotes the length of the time interval over which the norms are taken. As a consequence of our definitions, we then conclude that there exists a constant  $C$  such that one has the bound

$$\|\mathcal{P}\mathbf{1}_+ H\|_{\gamma, \eta; \varepsilon} \leq C(T + \varepsilon)^{\theta} \|H\|_{\delta, \eta_2; \varepsilon}.$$

Combining these remarks, we see that for every  $K > 1$  there exist a final time  $T$  and a constant  $\varepsilon_0$  such that, for all  $\varepsilon \in [0, \varepsilon_0]$ , the map  $\mathcal{M}$  defined in (4.21) maps the ball of radius  $K$  in  $\mathcal{D}_{\varepsilon}^{\gamma, \eta}$  into itself and is a contraction there, provided that the underlying model  $\Pi \in \mathcal{M}_{\varepsilon}$  satisfies  $|\Pi|_{\varepsilon} \leq K$  and that the initial condition  $h_0^{(\varepsilon)}$  satisfies  $\|h_0^{(\varepsilon)}\|_{\eta, \varepsilon} \leq K/(2C)$  for  $C$  as in (4.12).

We now turn to the second part of the statement, namely the question of convergence as  $\varepsilon \rightarrow 0$ . We denote by  $\mathcal{M}_T$  the fixed point map given in (4.21), where we make explicit the dependency on the terminal time  $T$ , and we write  $\mathcal{M}_T^{(\varepsilon)}$  for the same map, but with initial condition  $h_0^{(\varepsilon)} \in \mathcal{C}_{\varepsilon}^{\eta}$  and with respect to some model  $\Pi^{(\varepsilon)} \in \mathcal{M}_{\varepsilon}$ . Collecting all of the previously obtained estimates, we

see that for  $H \in \mathcal{D}_0^{\gamma, \eta}$  and  $H^{(\varepsilon)} \in \mathcal{D}_\varepsilon^{\gamma, \eta}$ , as well as corresponding models  $\Pi \in \mathcal{M}_0$  and  $\Pi^{(\varepsilon)} \in \mathcal{M}_\varepsilon$ , the fixed point map  $\mathcal{M}$  satisfies the bound

$$\begin{aligned} \|\mathcal{M}_T^{(\varepsilon)}(H^{(\varepsilon)}); \mathcal{M}_T(H)\|_\varepsilon &\lesssim (T + \varepsilon)^\theta \|H^{(\varepsilon)}; H\|_{\gamma, \eta; \varepsilon} \\ &\quad + \|\Pi^{(\varepsilon)}; \Pi\|_\varepsilon + \|h_0^{(\varepsilon)}; h_0\|_{\eta; \varepsilon}, \end{aligned}$$

where the proportionality constant is uniform over  $T, \varepsilon$  sufficiently small, as well as underlying models, initial conditions, and modelled distributions  $H, H^{(\varepsilon)}$  belonging to a ball of fixed radius in the corresponding ‘norms’. It immediately follows that for sufficiently small final time  $T$ , one has

$$\|H^{(\varepsilon)}; H\|_{\gamma, \eta; \varepsilon} \lesssim \|\Pi^{(\varepsilon)}; \Pi\|_\varepsilon + \|h_0^{(\varepsilon)}; h_0\|_{\eta; \varepsilon}. \quad (4.22)$$

It remains to show that if  $H$  is a solution to (4.21) up to some specified final time  $T$ , then the corresponding fixed point problem for  $\mathcal{M}_T^{(\varepsilon)}$  also has a solution up to the same time  $T$ , provided that  $\varepsilon$  is small enough, and the two underlying models and initial conditions are sufficiently close. This is not completely trivial since it may well happen that  $T$  is sufficiently large so that  $\mathcal{M}_T$  is no longer a contraction.

In view of (4.22), it suffices to obtain a bound on the solution, as well as the difference between solutions, at positive times in the same spaces  $\mathcal{C}_\varepsilon^\eta$  that we choose our initial condition in, so that we can iterate the bounds (4.22). (See also the construction of maximal solutions in [Hai14, Proposition 7.11] which shows that a restarted solution is again a solution of the original fixed point problem.) This on the other hand immediately follows from Proposition 4.8.  $\square$

To conclude this section, let us mention a straightforward way in which the solution map constructed in Theorem 4.16 actually relates to a PDE problem. Recall that, given any smooth (actually continuous is enough) function  $\zeta$ , the construction of Section 3.6 yields a family of maps  $\mathcal{L}_\varepsilon: \mathcal{C}^\infty \rightarrow \mathcal{M}$  lifting  $\zeta$  to an admissible model  $(\Pi, \Gamma) = \mathcal{L}_\varepsilon(\zeta)$ . The following result is then immediate:

**PROPOSITION 4.17.** Let  $h_0 \in \mathcal{C}^\gamma$  with  $\gamma$  as in Theorem 4.16 and, given  $\varepsilon \in \mathbf{R}$  and  $\zeta \in \mathcal{C}^0$ , let  $H \in \mathcal{D}_\varepsilon^{\gamma, \eta}$  be the local solution to (4.20) given by Theorem 4.16 for the restriction to  $\mathcal{T}$  of the canonical model  $\mathcal{L}_\varepsilon(\zeta)$ . Then, the function  $h = \mathcal{R}H$  is the classical (local) solution to the PDE

$$\partial_t h = \partial_x^2 h + \sum_{j=1}^m \varepsilon^{j-1} \hat{a}_j (\partial_x h)^{2j} + \zeta.$$

*Proof.* Applying the reconstruction operator to both sides of (4.20) and using the facts that the model  $\mathcal{L}_\varepsilon(\zeta)$  is admissible, that  $\mathcal{RP}\mathbf{1}_+ = P * \mathbf{1}_+ \mathcal{R}$

(see [Hai14, Section 4]), and that  $\mathcal{R}Q_{\leq 0}H = \mathcal{R}H$ , we see that

$$h = P * \mathbf{1}_+ \left( \zeta + \sum_{j=1}^m \hat{a}_j \mathcal{R}(\hat{\mathcal{E}}^{j-1}(Q_{\leq 0}(\mathcal{D}H)^{2j})) \right) + Ph_0,$$

where  $\mathbf{1}_+$  denotes the indicator function of the set  $\{t \geq 0\}$ . The claim now follows from the fact that the reconstruction operator obtain for the model  $\mathcal{L}_\varepsilon(\zeta)$  satisfies

$$\mathcal{R}(\hat{\mathcal{E}}^{j-1}(Q_{\leq 0}(\mathcal{D}H)^{2j})) = \varepsilon^{j-1}(\partial_x h)^{2j},$$

as a consequence of (3.35) which holds on  $\mathcal{T}$  by restriction.  $\square$

REMARK 4.18. Note that the parameter  $\varepsilon$  only enters in the construction of the model  $\mathcal{L}_\varepsilon(\zeta)$ . In particular, the solution map built in Theorem 4.16 does not itself have any knowledge of  $\varepsilon$ . This is the crucial feature of our construction that then allows us to send  $\varepsilon$  to 0 in a ‘transparent’ way.

## 5. Renormalization

The purpose of this section is to build a family of transformations on the space  $\mathcal{M}$  of all admissible models for the regularity structure  $(\mathcal{T}, \mathcal{G})$  (as opposed to  $(\mathcal{T}_{\text{ex}}, \mathcal{G})$  where we would not find any convergent renormalized model). These transformations will be of the type

$$\hat{\Pi}_x \tau = (\Pi_x \otimes f_x) \Delta^{\text{Wick}} M_0 \tau, \quad \hat{f}_x(\sigma) = f_x(\hat{M}^{\text{Wick}} \sigma), \quad (5.1)$$

where  $M_0: \mathcal{T} \rightarrow \mathcal{T}$ ,  $\hat{M}^{\text{Wick}}: \mathcal{T}_+ \rightarrow \mathcal{T}_+$ , and  $\Delta^{\text{Wick}}: \mathcal{T} \rightarrow \mathcal{T} \otimes \mathcal{T}_+$  are linear maps with additional properties guaranteeing that  $(\hat{\Pi}, \hat{f})$  is again an admissible model. Of course, we could also have just defined one single map instead of the composition  $\Delta^{\text{Wick}} M_0$ , but it turns out that the effects of the two factors are easier to analyse separately.

**5.1. Renormalization of the average speed.** We start by discussing the map  $M_0$  since this is easier to define. At the level of the equation, the effect of  $M_0$  will simply be to add a constant term to the right hand side. Denote by  $\mathcal{B} \subset \mathcal{T}$  the set of canonical basis vectors that are of one of the following two types:

$$\begin{aligned} \tau &= \mathcal{E}^\ell(\Psi^{2\ell} \mathcal{I}'(\mathcal{E}^m(\Psi^{2m+2})) \mathcal{I}'(\mathcal{E}^n(\Psi^{2n+2}))), \\ \tau &= \mathcal{E}^\ell(\Psi^{2\ell+1} \mathcal{I}'(\mathcal{E}^m(\Psi^{2m+1} \mathcal{I}'(\mathcal{E}^n(\Psi^{2n+2}))))), \end{aligned} \quad (5.2)$$

where  $\ell, m, n \geq 0$  are positive integers. Note that in both cases one has  $|\tau| = -2(\ell + m + n + 2)\kappa$ . For any  $\tau \in \mathcal{B}$ , we then define  $L_\tau: \mathcal{T} \rightarrow \mathcal{T}$  by setting  $L_\tau \tau = \mathbf{1}$  and  $L_\tau \bar{\tau} = 0$  for every canonical basis vector  $\bar{\tau} \neq \tau$ .

Finally, given constants  $C_\tau \in \mathbf{R}$ , we set

$$M_0 = \exp\left(-\sum_{\tau \in \mathcal{B}} C_\tau L_\tau\right) = 1 - \sum_{\tau \in \mathcal{B}} C_\tau L_\tau. \quad (5.3)$$

This defines a map  $(\Pi, f) \mapsto (\hat{\Pi}, \hat{f})$  on admissible models  $(\Pi, f) \in \mathcal{M}$  by  $\hat{\Pi}_z \tau = \Pi_z M_0 \tau$  and  $\hat{f}_z = f_z$ , taking reconstruction operator  $\mathcal{R}$  associated to  $(\Pi, f)$  to  $\hat{\mathcal{R}}$  associated to  $(\hat{\Pi}, \hat{f})$ . In this case, since  $\hat{f} = f$ , the spaces  $\mathcal{D}^\gamma$  of modelled distributions are the same for the two models. This transformation enjoys the following properties:

- PROPOSITION 5.1. (1) For every  $\Gamma \in \mathcal{G}$  and  $\tau \in \mathcal{T}$ ,  $M_0 \Gamma \tau = \Gamma M_0 \tau$ ;  
 (2)  $M_0 \mathcal{I}'(\tau) = \mathcal{I}'(\tau)$ ;  
 (3) the map  $(\Pi, f) \mapsto (\hat{\Pi}, \hat{f})$  is continuous on the space of all models for  $(\mathcal{T}, \mathcal{G})$  and maps the space  $\mathcal{M}$  of admissible models into itself;  
 (4) for every  $H \in \mathcal{D}^\gamma$  with  $\gamma > 0$  such that the coefficient  $u_\tau$  of  $H$  is constant for  $\tau \in \mathcal{B}$ , one has the identity  $\hat{\mathcal{R}}H = \mathcal{R}H - \sum_{\tau \in \mathcal{B}} C_\tau u_\tau$ .

*Proof.* Note first that  $\Delta L_\tau \tau = \mathbf{1} \otimes \mathbf{1} = (L_\tau \otimes \mathbf{1}) \Delta \tau$ . Furthermore, for any  $\bar{\tau} \in \mathcal{T}$ , one has  $\Delta \bar{\tau} = \bar{\tau} \otimes \mathbf{1} + \sum \bar{\tau}^{(1)} \otimes \bar{\tau}^{(2)}$  with  $|\tau^{(1)}| < |\tau|$  and by checking the few cases of  $\bar{\tau} \in \mathcal{T}$  with  $|\bar{\tau}| > 0$  we see that  $\bar{\tau}^{(1)} \notin \mathcal{B}$  for any  $\bar{\tau} \in \mathcal{T}$ . It immediately follows that if  $\bar{\tau} \neq \tau$ , one has  $(L_\tau \otimes \mathbf{1}) \Delta \bar{\tau} = 0$ , thus concluding the proof of (1). (2) follows from the definition of  $M_0$  since  $\mathcal{I}'(\tau) \notin \mathcal{B}$ . (3) follows from 1 together with [Hai14, Proposition 2.30]. The last property then follows from (5.3) and the definition of  $\mathcal{R}$ .  $\square$

**5.2. Wick renormalization.** We now describe maps  $\hat{M}^{\text{Wick}}: \mathcal{T}_+ \rightarrow \mathcal{T}_+$ , and  $\Delta^{\text{Wick}}: \mathcal{T} \rightarrow \mathcal{T} \otimes \mathcal{T}_+$  corresponding to *Wick renormalization* with respect to the Gaussian structure generated by solutions to the linearized equation. Here the extended regularity structure  $\mathcal{T}_{\text{ex}}$  is particularly useful. The way the maps  $\hat{M}^{\text{Wick}}$  and  $\Delta^{\text{Wick}}$  are constructed is to first build them on  $\mathcal{T}_{\text{ex}}$  and then define them on  $\mathcal{T}$  simply by restriction. The key defining properties on the renormalization group, that  $(\hat{\Pi}, \hat{f})$  defined through (5.1) is in  $\mathcal{M}$  and that  $\Delta^{\text{Wick}} \tau = \tau \otimes \mathbf{1} + \sum \hat{\tau}^{(1)} \otimes \hat{\tau}^{(2)}$ , with  $|\hat{\tau}^{(1)}| > \tau$ , are inherited by descent from  $\mathcal{T}_{\text{ex}}$ , since  $\mathcal{T}$  is a sector of  $\mathcal{T}_{\text{ex}}$ . Hence it suffices to construct  $\hat{M}^{\text{Wick}}: \mathcal{T}_+ \rightarrow \mathcal{T}_+$ , and  $\Delta^{\text{Wick}}: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}} \otimes \mathcal{T}_+$ .

We first build an associated map  $M^{\text{Wick}}: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}}$  depending on a parameter  $C^{\text{W}} \in \mathbf{R}$  by setting

$$M^{\text{Wick}} = \exp(-C^{\text{W}} L^{\text{Wick}}), \quad (5.4)$$



where the generator  $L^{\text{Wick}}$  iterates over every occurrence of the subexpression  $\Psi^2$  and sends it to **1**. More formally,

$$L^{\text{Wick}} \mathcal{E} = L^{\text{Wick}} \mathbf{1} = 0, \quad L^{\text{Wick}} \Psi^j = \binom{j}{2} \Psi^{j-2}, \quad (5.5a)$$

for every  $j \geq 2$ . This is extended to  $\mathcal{T}_{\text{ex}}$  by imposing the *Leibniz rule*,

$$L^{\text{Wick}}(\tau \mathcal{I}'(\bar{\tau})) = L^{\text{Wick}}(\tau) \mathcal{I}'(\bar{\tau}) + \tau \mathcal{I}'(L^{\text{Wick}} \bar{\tau}), \quad (5.5b)$$

as well as the commutation relations

$$\begin{aligned} L^{\text{Wick}} \mathcal{I}'(\tau) &= \mathcal{I}'(L^{\text{Wick}} \tau), & L^{\text{Wick}} \mathcal{E}^\ell(\tau) &= \mathcal{E}^\ell(L^{\text{Wick}} \tau), \\ L^{\text{Wick}}(X^\ell \tau) &= X^\ell(L^{\text{Wick}} \tau), \end{aligned} \quad (5.5c)$$

for any two formal expressions  $\tau$  and  $\bar{\tau}$  with  $\bar{\tau} \neq \mathcal{E}$ . Since all elements of  $\mathcal{T}_{\text{ex}}$  can be obtained in this way, this defines  $L^{\text{Wick}}$  uniquely. In particular, these definitions imply that

$$M^{\text{Wick}} \Psi^m = H_m(\Psi, C^{\text{W}}), \quad (5.6)$$

where  $H_m(x, c)$  denote the generalized Hermite polynomials given by  $H_2(x, c) = x^2 - c$ ,  $H_4(x, c) = x^4 - 6cx^2 + 3c^2$ , and so on.

Denote now by  $\mathfrak{R}_0$  the set of all linear maps  $M: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}}$  which fix  $\mathcal{E}$  and **1** and commute with the abstract integration operators  $\mathcal{I}$ ,  $\mathcal{I}'$  and  $\mathcal{E}^\ell$ . Recall then from [Hai14, Section 8] that if  $M \in \mathfrak{R}_0$ , then one can uniquely associate to it maps  $\Delta^M: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}} \otimes \mathcal{T}_+$  and  $\hat{M}: \mathcal{T}_+ \rightarrow \mathcal{T}_+$  satisfying the properties

$$\begin{aligned} \hat{M} \mathcal{J}_k &= \mathcal{M}(\mathcal{J}_k \otimes 1) \Delta^M, \\ \hat{M} \mathcal{E}_k^\ell &= \mathcal{M}(\mathcal{E}_k^\ell \otimes 1) \Delta^M, \\ (1 \otimes \mathcal{M})(\Delta \otimes 1) \Delta^M &= (M \otimes \hat{M}) \Delta, \\ \hat{M}(\sigma_1 \sigma_2) &= (\hat{M} \sigma_1)(\hat{M} \sigma_2), \quad \hat{M} X^k = X^k, \end{aligned} \quad (5.7)$$

where  $\mathcal{M}: \mathcal{T}_+ \otimes \mathcal{T}_+ \rightarrow \mathcal{T}_+$  denotes the product in the Hopf algebra  $\mathcal{T}_+$ .

REMARK 5.2. At first sight, our regularity structure appears not to be exactly of the type considered in [Hai14, Section 8]. However, it follows from (3.9b) that  $\mathcal{E}^\ell$  is nothing but an abstract integration map of order  $\ell$  on  $\mathcal{T}_{\text{ex}}$ . It is then straightforward to verify that the results of that section still apply *mutatis mutandis* to the present situation.

We then define the renormalization group  $\mathfrak{R}$  for  $\mathcal{T}_{\text{ex}}$  as follows:

DEFINITION 5.3. A linear map  $M \in \mathfrak{R}_0$  belongs to  $\mathfrak{R}$  if the associated map  $\Delta^M$  is such that  $\Delta^M \tau = \tau \otimes \mathbf{1} + \sum \tau_M^{(1)} \otimes \tau_M^{(2)}$ , for some elements  $\tau_M^{(i)}$  satisfying  $|\tau_M^{(1)}| > |\tau|$ .

REMARK 5.4. The definition of  $\mathfrak{R}$  given here does not appear to match the definition given in [Hai14, Definition 8.41], where we also imposed a similar condition on a second operator  $\hat{\Delta}^M$  built from  $M$ . It turns out however that Definition 5.3 actually implies that second condition, as we show in Appendix B.

With these definitions at hand, given  $M \in \mathfrak{R}$ , we can use it to build a map  $(\Pi, f) \mapsto (\Pi^M, f^M)$  mapping admissible models to admissible models by setting

$$\Pi_z^M = (\Pi_z \otimes f_z) \Delta^M, \quad f_z^M = f_z \circ \hat{M},$$

see [Hai14, Theorem 8.44]. It is furthermore straightforward to verify that if an admissible model  $(\Pi, f)$  consists of smooth functions satisfying the identity (3.26) then, as a consequence of the second identity in (5.7), the renormalized model  $(\Pi^M, f^M)$  is also guaranteed to satisfy this identity. The remainder of this section is devoted to the proof that the map  $M^{\text{Wick}}$  given in (5.4) does indeed belong to  $\mathfrak{R}$ . In order to do this, we first make a few general considerations. Given a linear map  $M: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}}$  in  $\mathfrak{R}_0$ , we first show the following result.

PROPOSITION 5.5. Let  $M \in \mathfrak{R}_0$  and let  $\Delta^M$  and  $\hat{M}$  be the unique maps satisfying (5.7). Let  $\tau$  be a canonical basis element of  $\mathcal{T}_{\text{ex}}$ , and let  $\Delta^M \tau = \tau_M^{(1)} \otimes \tau_M^{(2)}$  (with summation implicit) be such that  $|\tau_M^{(1)}| \geq |\tau|$ . Then, one has

$$\begin{aligned} \Delta^M \mathcal{E}^\ell(\tau) &= (\mathcal{E}^\ell \otimes 1) \Delta^M \tau - \sum_{|k| > |\tau| + \ell} \frac{X^k}{k!} \otimes \mathcal{E}_k^\ell(\tau_M^{(1)}) \tau_M^{(2)}, \\ \Delta^M \mathcal{I}'(\tau) &= (\mathcal{I}' \otimes 1) \Delta^M \tau - \sum_{|k| > |\tau| + 1} \frac{X^k}{k!} \otimes \mathcal{I}_{k+1}'(\tau_M^{(1)}) \tau_M^{(2)}, \end{aligned}$$

and similarly for  $\Delta^M \mathcal{I}(\tau)$ .

*Proof.* We use the shorthand  $D = (1 \otimes \mathcal{M})(\Delta \otimes 1)$ . We only give a proof for  $\mathcal{I}(\tau)$ . The proofs for  $\mathcal{I}'(\tau)$  and  $\mathcal{E}^\ell(\tau)$  are identical since these operators have exactly the same algebraic properties. Combining (3.9b) with the first identity in (5.7)

and the fact that  $\mathcal{I}$  and  $M$  commute by assumption, we obtain the identity

$$\begin{aligned} (M \otimes \hat{M})\Delta\mathcal{I}(\tau) &= (\mathcal{I}M \otimes \hat{M})\Delta\tau + \sum_{|k+\ell|<|\tau|+1} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \hat{M}\mathcal{J}_{k+\ell+1}(\tau) \\ &= (\mathcal{I} \otimes 1)D\Delta^M\tau + \sum_{|k+\ell|<|\tau|+1} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \mathcal{J}_{k+\ell+1}(\tau_M^{(1)})\tau_M^{(2)}. \end{aligned}$$

On the other hand, using again (3.9b), we also have the identity

$$\begin{aligned} D(\mathcal{I} \otimes 1)\Delta^M\tau &= (\mathcal{I} \otimes 1)D\Delta^M\tau + \sum_{k,\ell} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \mathcal{M}(\mathcal{J}_{k+\ell+1} \otimes 1)\Delta^M\tau \\ &= (\mathcal{I} \otimes 1)D\Delta^M\tau + \sum_{|k+\ell|<|\tau_M^{(1)}|+1} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \mathcal{J}_{k+\ell+1}(\tau_M^{(1)})\tau_M^{(2)}, \end{aligned}$$

so that, since  $|\tau_M^{(1)}| \geq |\tau|$  by assumption, one has

$$D(\mathcal{I} \otimes 1)\Delta^M\tau = (M \otimes \hat{M})\Delta\mathcal{I}(\tau) + \sum_{|k+\ell|>|\tau|+1} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \mathcal{J}_{k+\ell+1}(\tau_M^{(1)})\tau_M^{(2)}. \quad (5.8)$$

At this stage we note that, if  $\{\tau_k\}$  is any collection of elements of  $\mathcal{T}_{\text{ex}}$  indexed by the multiindex  $k$ , then it follows from the action of  $\Delta$  on  $X^m$  that one has the identity

$$D\left(\frac{X^m}{m!} \otimes \tau_m\right) = \sum_{k+\ell=m} \frac{X^k}{k!} \otimes \frac{X^\ell}{\ell!} \tau_m.$$

Combining this with (5.8), we conclude that

$$D(\mathcal{I} \otimes 1)\Delta^M\tau = (M \otimes \hat{M})\Delta\mathcal{I}(\tau) + D \sum_{|k|>|\tau|+1} \frac{X^k}{k!} \otimes \mathcal{J}_{k+1}(\tau_M^{(1)})\tau_M^{(2)}.$$

Since furthermore  $(M \otimes \hat{M})\Delta\mathcal{I}(\tau) = D\Delta^M\mathcal{I}\tau$  by the definition (5.7) of  $\Delta^M$  and since the linear map  $D$  is invertible (it differs from the identity by a nilpotent operator), the claim follows at once.  $\square$

**PROPOSITION 5.6.** Let  $M \in \mathfrak{R}_0$ , let  $k \geq 0$  and let  $V_0, \dots, V_k$  be sectors of  $\mathcal{T}_{\text{ex}}$  such that, if  $\tau_i \in V_i$ , then  $\tau_0 \cdots \tau_k \in \mathcal{T}_{\text{ex}}$  and  $M(\tau_0 \cdots \tau_k) = (M\tau_0) \cdots (M\tau_k)$ . Then, one also has  $\Delta^M(\tau_0 \cdots \tau_k) = (\Delta^M\tau_0) \cdots (\Delta^M\tau_k)$ .

*Proof.* Let  $\tau_i \in V_i$  as in the statement and set  $\tau = \tau_0 \cdots \tau_k$ . Since  $\hat{M}$  is a multiplicative morphism, it follows from our assumption that

$$(M \otimes \hat{M})\Delta\tau = \prod_{i=0}^k (M \otimes \hat{M})\Delta\tau_i. \quad (5.9)$$

Since  $\Delta^M\tau = D^{-1}(M \otimes \hat{M})\Delta\tau$  (with  $D$  as above) and since  $D$  is a multiplicative morphism, the claim follows at once by applying  $D^{-1}$  to both sides of (5.9).  $\square$

We then have

PROPOSITION 5.7. Let  $M^{\text{Wick}}$  be as above, let  $\Delta^{\text{Wick}}$  and  $\hat{M}^{\text{Wick}}$  be the corresponding maps satisfying (5.7), and let  $\tau \in \mathcal{T}_{\text{ex}}$  be a canonical basis vector of the form

$$\tau = \Psi^m \prod_{i=1}^k \mathcal{I}'(\tau_i), \quad (5.10)$$

where  $k, m \geq 0$ , and the  $\tau_i$  are canonical basis vectors with  $\tau_i \neq \mathcal{E}$ . Then, one has

$$\Delta^{\text{Wick}}\tau = (M^{\text{Wick}}\Psi^m \otimes \mathbf{1}) \prod_{i=1}^k \Delta^{\text{Wick}}\mathcal{I}'(\tau_i). \quad (5.11)$$

*Proof.* We first note the following very important fact. By the construction of  $\mathcal{T}_{\text{ex}}$ , if  $\mathcal{I}'(\tau) \in \mathcal{T}_{\text{ex}}$  with  $\tau \neq \mathcal{E}$ , then  $\tau$  cannot contain any factor  $\mathcal{E}$  by the construction of  $\mathcal{T}_{\text{ex}}$ . Therefore, by construction,  $L^{\text{Wick}}\tau_i$  does not contain any summand proportional to  $\mathcal{E}$  either. As a consequence of the ‘Leibnitz rule’ satisfied by the  $L_j$ , this then shows that, for every  $p \geq 0$ ,

$$(L^{\text{Wick}})^p\tau = \sum_{p_0+\cdots+p_k=p} ((L^{\text{Wick}})^{p_0}\Psi^m) \prod_{i=1}^k \mathcal{I}'((L^{\text{Wick}})^{p_i}\tau_i),$$

which in particular implies that

$$M^{\text{Wick}}\tau = (M^{\text{Wick}}\Psi^m) \prod_{i=1}^k \mathcal{I}'(M^{\text{Wick}}\tau_i). \quad (5.12)$$

Similarly, one verifies that if one writes  $\Delta\tau_i = \tau_i^{(1)} \otimes \tau_i^{(2)}$  (with an implicit summation over such terms), then none of the terms  $\tau_i^{(1)}$  can be equal to  $\mathcal{E}$ . Applying the definition of  $\Delta$ , one also verifies that the linear span of the vectors  $\Psi^m$  is stable under the action of the structure group  $\mathcal{G}$ . Combining these

observations, we see that Proposition 5.6 applies, so that

$$\Delta^{\text{Wick}} \tau = (\Delta^{\text{Wick}} \Psi^m) \prod_{i=1}^k \Delta^{\text{Wick}} \mathcal{I}'(\tau_i).$$

The fact that  $\Delta^{\text{Wick}} \Psi^m = (M^{\text{Wick}} \Psi^m \otimes \mathbf{1})$  can easily be verified ‘by hand’ from (5.7).  $\square$

As a corollary of these two results, it is now easy to show that  $M^{\text{Wick}} \in \mathfrak{R}$ .

COROLLARY 5.8. One has  $M^{\text{Wick}} \in \mathfrak{R}$ .

*Proof.* As a consequence of the construction of  $\mathcal{T}_{\text{ex}}$  given in Section 3, we see that every one of its basis elements can be built from  $\mathcal{E}$  by making use of the operations  $\tau \mapsto \mathcal{I}(\tau)$ ,  $\tau \mapsto \mathcal{I}'(\tau)$ ,  $\tau \mapsto \mathcal{E}^\ell(\tau)$ ,  $\tau \mapsto X^\ell \tau$ , as well as  $(\tau_1, \dots, \tau_k) \mapsto \Psi^m \prod_{i=1}^k \mathcal{I}'(\tau_i)$  with  $\tau_i \neq \mathcal{E}$ . Since  $\Delta^{\text{Wick}} \mathcal{E} = \mathcal{E} \otimes \mathbf{1}$  and since the upper triangular structure of  $\Delta^{\text{Wick}}$  is preserved under all of these operations by Propositions 5.5 and 5.7, the claim follows.  $\square$

**5.3. Renormalized equations.** Let now  $(\Pi, f) = \mathcal{L}_\varepsilon(\zeta)$ , where  $\zeta$  is a continuous function and the canonical lift  $\mathcal{L}_\varepsilon$  is as in Section 3.6. We furthermore consider the renormalized model  $(\hat{\Pi}, \hat{f})$  given by (5.1) with  $M_0$  and  $M^{\text{Wick}}$  as in (5.3) and (5.4). In particular,  $M^{\text{Wick}}$  depends on the renormalization constant  $C^{\text{W}}$  while  $M_0$  depends on a collection of renormalization constants  $C_\tau$ .

The aim of this section is to show that if  $H$  solves the abstract fixed point problem (4.20) for the model  $(\hat{\Pi}, \hat{f})$ , then  $h = \hat{\mathcal{R}}H$ , where  $\hat{\mathcal{R}}$  the reconstruction operator associated to the renormalized model, can be identified with the solution to a modified PDE. In order to derive this new equation, we combine the explicit abstract form of the solutions with the product formula given by Proposition 5.7. The result is the following, where  $\mathcal{C}$  denotes the space of continuous functions on  $\mathbf{R} \times S^1$ :

PROPOSITION 5.9. Let  $h_0 \in \mathcal{C}^1$  and, given  $\varepsilon \in \mathbf{R}$  and  $\zeta \in \mathcal{C}^0$ , let  $H \in \mathcal{D}_{\varepsilon}^{\gamma, \eta} \subset \mathcal{D}^{\gamma, \eta}$  be the local solution to (4.20) given by Theorem 4.16 for the renormalized model  $(\hat{\Pi}, \hat{f})$  obtained from  $\mathcal{L}_\varepsilon(\zeta)$  in the way described above. Then, there exists a constant  $c$  such that the function  $h = \hat{\mathcal{R}}H$  is the classical (local) solution to the PDE

$$\partial_t h = \partial_x^2 h + \sum_{j=1}^m \varepsilon^{j-1} \hat{a}_j H_{2j}(\partial_x h, C^{\text{W}}) + c + \zeta,$$

with initial condition  $h_0$ .

REMARK 5.10. The constant  $c$  is a suitable linear combination of the constants  $C_\tau$  appearing in the definition (5.3) of  $M_0$ , with coefficients depending on the constants  $\hat{a}_j$ . In principle, one can derive an explicit expression for it, but this expression does not seem to be of particular interest. The only important fact is that if we write

$$\tau_1 = \Psi \mathcal{I}'(\Psi \mathcal{I}'(\Psi^2)), \quad \tau_2 = \mathcal{I}'(\Psi^2)^2,$$

then the corresponding renormalization constants  $c_{\tau_1}$  and  $c_{\tau_2}$  only ever arise as a multiple of  $4c_{\tau_1} + c_{\tau_2}$ . This is important since, as we will see in Theorem 6.5 below, these renormalization constants need to be chosen to diverge logarithmically as  $\varepsilon \rightarrow 0$  and the particular form of this linear combination guarantees that these logarithmic divergencies cancel out and are therefore not visible in the renormalized equations.

*Proof.* As in the proof of Proposition 4.17, we use the fact that the renormalized model is admissible to conclude that, when applying  $\hat{\mathcal{R}}$  to both sides of (4.20), the function  $h = \hat{\mathcal{R}}H$  satisfies the identity

$$h = P * \mathbf{1}_+ \left( \zeta + \sum_{j=1}^m \hat{a}_j \hat{\mathcal{R}}(\hat{\mathcal{E}}^{j-1}(\mathcal{Q}_{\leq 0}(\mathcal{D}H)^{2j})) \right) + Ph_0. \quad (5.13)$$

At this stage, the proofs diverge since it is no longer the case that  $\hat{\mathcal{R}}$  preserves the usual product. The only fact that we can use is that  $(\hat{\mathcal{R}}F)(z) = (\hat{\Pi}_z F(z))(z)$  (as a consequence of the assumption that our models are built from a continuous function  $\zeta$ ), combined with the definition of the renormalized model  $\hat{\Pi}$ .

Denoting by  $\mathcal{R}^{\text{Wick}}$  the reconstruction operator associated to the model  $(\Pi_x \otimes f_x) \Delta^{\text{Wick}}$ , then it follows from (3.23) and (5.1) that one has the identity

$$\hat{\mathcal{R}}U = \mathcal{R}^{\text{Wick}} M_0 U. \quad (5.14)$$

Furthermore, as a consequence of the second identity in (5.7) combined with (3.26), one has the identity

$$(\mathcal{R}^{\text{Wick}} \hat{\mathcal{E}}^\ell(U))(z) = \varepsilon^\ell (\mathcal{R}^{\text{Wick}} U)(z), \quad (5.15)$$

provided that the underlying model  $(\Pi, f)$  is of the form  $\mathcal{L}_\varepsilon(\zeta)$  for a continuous  $\zeta$ . Indeed, by Remark 3.18, it suffices to verify that the Wick renormalized model satisfies (3.26), but this follows from the fact that this identity is satisfied by the canonical lift  $\mathcal{L}_\varepsilon(\zeta)$ , combined with the second identity in (5.5c) as well as the definition (5.7) of  $\hat{M}^{\text{Wick}}$ . Note though that (5.15) fails in general if we replace

$\mathcal{R}^{\text{Wick}}$  by  $\hat{\mathcal{R}}$ . Take for example  $\ell = 1$  and  $U = \Psi^4$ , and set  $\psi(z) = (K * \zeta)(z)$ . Then, one has

$$(\hat{\mathcal{R}}\hat{\mathcal{E}}(U))(z) = \varepsilon H_4(\psi(z), C^{\text{W}}) - C_\tau \neq \varepsilon H_4(\psi(z), C^{\text{W}}) = \varepsilon(\hat{\mathcal{R}}U)(z).$$

It follows from the fact that  $\mathcal{DPF}$  differs from  $\mathcal{IF}$  by a Taylor polynomial at each point that if  $H$  is the solution to (4.20), then one can write

$$\mathcal{D}H(z) = \Psi + U(z),$$

where the remainder  $U$  only contains components proportional to either  $\mathbf{1}$ ,  $X$ , or  $\mathcal{I}'(\tau)$  with  $\tau \neq \mathcal{E}$ . In particular, none of the components belongs to  $\mathcal{B}$ , so that one has the identity

$$(\hat{\mathcal{R}}\mathcal{D}H)(z) = (\mathcal{R}^{\text{Wick}}\mathcal{D}H)(z) = (\Pi_z\Psi)(z) + ((\Pi_z \otimes f_z)\Delta^{\text{Wick}}U(z))(z).$$

On the other hand, for  $\ell \geq 0$ , we can apply the reconstruction operator to  $\hat{\mathcal{E}}^\ell((\mathcal{D}H)^{2\ell+2})$  and combine (5.14) with (5.15) and the definition of  $M_0$  thus yielding

$$\begin{aligned} (\hat{\mathcal{R}}\hat{\mathcal{E}}^\ell(Q_{\leq 0}(\mathcal{D}H)^{2\ell+2}))(z) &= (\mathcal{R}^{\text{Wick}}\hat{\mathcal{E}}^\ell(Q_{\leq 0}(\mathcal{D}H)^{2\ell+2}))(z) + c \\ &= \varepsilon^\ell(\mathcal{R}^{\text{Wick}}(\mathcal{D}H)^{2\ell+2})(z) + c, \end{aligned} \quad (5.16)$$

for some constant  $c$ . It thus remains to compute  $\mathcal{R}^{\text{Wick}}(\mathcal{D}H)^m$  for arbitrary  $m$ . As a consequence of Proposition 5.7 and (5.6), we have

$$\begin{aligned} \Delta^{\text{Wick}}(\mathcal{D}H(z))^m &= \sum_{k+\ell=m} \binom{m}{k} (M^{\text{Wick}}\Psi^k \otimes \mathbf{1})(\Delta^{\text{Wick}}U(z))^\ell \\ &= \sum_{k+\ell=m} \binom{m}{k} (H_k(\Psi, C^{\text{W}}) \otimes \mathbf{1})(\Delta^{\text{Wick}}U(z))^\ell. \end{aligned}$$

At this stage, we use the fact that since our original model originates from a canonical lift by assumption, it has the property that  $\Pi_x\tau\bar{\tau} = \Pi_x\tau\Pi_x\bar{\tau}$ . Applying  $\Pi_z \otimes f_z$  to both sides of this equality and combining this with the fact that  $f_z$  is also multiplicative, we conclude that

$$\begin{aligned} (\mathcal{R}^{\text{Wick}}(\mathcal{D}H(z))^m)(z) &= \sum_{k+\ell=m} \binom{m}{k} H_k((\Pi_z\Psi)(z), C^{\text{W}}) \\ &\quad \times ((\Pi_z \otimes f_z)\Delta^{\text{Wick}}U(z))(z)^\ell \\ &= H_m((\mathcal{R}^{\text{Wick}}\mathcal{D}H)(z), C^{\text{W}}). \end{aligned}$$

Combining this with (5.16) and (5.13), the claim follows.  $\square$

## 6. Convergence of the models

In this section, we now show how the renormalization maps from the previous section can be used to renormalize the models built from regularizations of space–time white noise. From now on, we will use a graphical shorthand notation similar to the one used in [Hai13] for symbols  $\tau \in \mathcal{W}$  which do not contain the symbol  $\mathcal{E}$ : dots represent the symbol  $\mathcal{E}$ , lines denote the operator  $\mathcal{I}'$ , and the joining of symbols by their roots denotes their product. For example, one has  $\dagger = \mathcal{I}'(\mathcal{E}) = \Psi$ ,  $\mathfrak{V} = \Psi^2$ ,  $\mathfrak{V}_\bullet = \Psi \mathcal{I}'(\Psi^2)$ , and so on. We will also assume from now on that  $(\mathcal{T}, \mathcal{G})$  has been truncated in the way specified in the beginning of Section 3.3.

With the same graphical notations as in (6.1), we also define the renormalization constants

$$C_0^{(\varepsilon)} = \text{diagram}, \quad C_2^{(\varepsilon)} = \text{diagram}, \quad C_3^{(\varepsilon)} = \text{diagram}, \quad (6.1)$$

where a plain arrow represents the kernel  $K'$ . We will see in Section 6.3 below that the last two constants diverge logarithmically as  $\varepsilon \rightarrow 0$ , but this is not important at the moment. Note also that  $C_0^{(\varepsilon)}$  is the left hand side of (1.13) and, for all  $\tau \in \mathcal{B} \setminus \{\mathfrak{V}_\bullet, \mathfrak{V}_\bullet\}$  defined in the prelude to (5.2), we set

$$C_\tau^{(\varepsilon)} = \mathbf{E}(\Pi^{(\varepsilon)} M^{\text{Wick}}_\tau)(0), \quad (6.2)$$

where  $M^{\text{Wick}}$  is the map defined in (5.4) with  $C^{\text{W}} = C_0^{(\varepsilon)}$  and  $\Pi^{(\varepsilon)}: \mathcal{T} \rightarrow \mathcal{C}$  denotes the linear map defined recursively by  $\Pi^{(\varepsilon)} \mathcal{E} = \xi^{(\varepsilon)}$  and

$$\Pi^{(\varepsilon)} \mathcal{E}^k(\tau) = \varepsilon^k \Pi^{(\varepsilon)} \tau, \quad \Pi^{(\varepsilon)} \mathcal{I}'(\tau) = K' * \Pi^{(\varepsilon)} \tau, \quad \Pi^{(\varepsilon)} \tau \bar{\tau} = (\Pi^{(\varepsilon)} \tau)(\Pi^{(\varepsilon)} \bar{\tau}).$$

Note that the functions  $\Pi^{(\varepsilon)} \tau$  are stationary, so the choice of the evaluation at 0 in (6.2) is irrelevant.

We then define a map  $M^{(\varepsilon)}$  acting on the space of admissible models by

$$M^{(\varepsilon)}: (\Pi, f) \mapsto (\hat{\Pi}, \hat{f}), \quad (6.3)$$

with  $(\hat{\Pi}, \hat{f})$  as in (5.1), where we set

$$C_{\mathfrak{V}_\bullet} = 2C_2^{(\varepsilon)}, \quad C_{\mathfrak{V}_\bullet} = 2C_3^{(\varepsilon)},$$

as well as  $C_\tau = C_\tau^{(\varepsilon)}$  for  $\tau \in \mathcal{B} \setminus \{\mathfrak{V}_\bullet, \mathfrak{V}_\bullet\}$ . With these notations at hand, the following is then the main result of this section.



THEOREM 6.1. Let  $\xi^{(\varepsilon)}$  be as in (1.11) and consider the sequence of models on  $\mathcal{T}$  given by

$$\mathfrak{M}_\varepsilon = M^{(\varepsilon)} \mathcal{L}_\varepsilon(\xi^{(\varepsilon)}).$$

Then, there exists a random model  $\mathfrak{M}$  such that  $|\mathfrak{M}_\varepsilon; \mathfrak{M}|_\varepsilon \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$ . Furthermore, the limiting model  $\mathfrak{M} = (\hat{\Pi}, \hat{f})$  is independent of the choice of mollifier  $\varrho$  and it satisfies  $\hat{\Pi}_\tau \tau = 0$  for every symbol  $\tau$  containing at least one occurrence of  $\mathcal{E}$ .

Before we turn to the proof of Theorem 6.1, we give a criterion allowing to verify whether a sequence of models converges in  $\mathcal{M}_\varepsilon$ .

**6.1. A convergence criterion.** The following result is very useful. Here, we fix a sufficiently regular wavelet basis/multiresolution analysis with compactly supported elements and we reuse the notation of [Hai14, Section 3.1]. In particular,  $\Psi$  is a finite set of functions in  $\mathcal{B}$  such that the wavelet basis is obtained by translations and rescalings of elements in  $\Psi$  (we use the notation  $\Psi$  to be consistent with [Hai14]. It should not be confused with the shorthand for  $\mathcal{I}'(\mathcal{E})$  used elsewhere in the paper). Here, we follow the usual convention, so  $\psi_z^n$  denotes a wavelet basis function at level  $n$  (scale  $2^{-n}$ ) centred at some point  $z$  in the level  $n$  dyadic set  $\Lambda^n$ . We normalize these basis functions so that their  $L^2$  norm (not the  $L^1$  norm as before!) equals 1.

Recall also that our definition of the spaces  $\mathcal{M}_\varepsilon$  involves the constant  $\bar{\gamma} = 1 - \frac{1}{32m}$  as defined in (4.7b).

PROPOSITION 6.2. Let  $(\Pi^{(\varepsilon)}, f^{(\varepsilon)})$  be a family of models for the regularity structure  $(\mathcal{T}, \mathcal{G})$  converging to a limiting model  $(\Pi, f)$  in the sense that  $\lim_{\varepsilon \rightarrow 0} |\Pi^{(\varepsilon)}; \Pi| = 0$ . Assume that, for some  $\delta > 0$ , one has

$$|f_z^{(\varepsilon)}(\mathcal{G}_\ell^k(\tau))| \leq C \varepsilon^{|\tau|+k-|\ell|+\delta}, \quad (6.4)$$

for  $\tau$  and  $k, \ell$  as in (4.7a), and that furthermore for  $\tau \in \mathcal{U}'$

$$|(\Pi_z^{(\varepsilon)} \tau)(\psi_z^n)| \leq C 2^{-(3n/2)-\bar{\gamma}n} \varepsilon^{|\tau|-\bar{\gamma}+\delta}, \quad (6.5)$$

for every  $n \geq 0$ , every  $z \in \Lambda^n$ , and every  $\psi \in \Psi$ . Then, one has  $\lim_{\varepsilon \rightarrow 0} |\Pi^{(\varepsilon)}; \Pi|_{\varepsilon,0} = 0$ .

*Proof.* We only need to show that, for any test function  $\eta \in \mathcal{B}$  with  $\int \eta(z) dz = 0$ , one has

$$|(\Pi_z^{(\varepsilon)} \tau)(\eta_z^\lambda)| \lesssim \lambda^{\bar{\gamma}} \varepsilon^{|\tau|-\bar{\gamma}+\delta},$$

provided that  $\lambda \leq \varepsilon$ , since this will then guarantee that  $\|\Pi^{(\varepsilon)}\|_\varepsilon \lesssim \varepsilon^\delta$ . We fix  $N \geq 0$  such that  $2^{-N} \leq \varepsilon \leq 2^{1-N}$  and we write

$$\eta_z^\lambda = \sum_{z' \in \Lambda^N} A_{z'}^N \varphi_{z'}^N + \sum_{n \geq N} \sum_{\psi \in \Psi} A_{z'}^{n, \psi} \psi_{z'}^n. \quad (6.6)$$

It is then a simple consequence of the scaling properties of these objects that one has the bounds

$$|A_{z'}^N| \lesssim 2^{3N/2} (\lambda/2^{-N}), \quad |A_{z'}^{n, \psi}| \lesssim \begin{cases} 2^{3n/2} (\lambda/2^{-n}) & \text{if } 2^{-n} \geq \lambda, \\ 2^{3n/2} (2^{-n}/\lambda)^3 & \text{otherwise.} \end{cases}$$

Note here that the factor  $2^{3n/2}$  comes from the fact that the functions  $\psi_z^n$  and  $\varphi_z^n$  appearing in (6.6) are normalized in  $L^2$  rather than in  $L^1$ . Furthermore, the factor  $\lambda/2^{-n}$  appearing in the first two bounds is a consequence of the fact that  $\eta$  integrates to 0 by assumption and the wavelet basis is sufficiently regular ( $C^2$  is enough).

We furthermore obtain from (6.5) and the fact that  $\Pi^{(\varepsilon)}$  converges to a limit (and therefore is bounded in  $\mathcal{M}$ , uniformly in  $\varepsilon$ ) the bound

$$\begin{aligned} |(\Pi_z^{(\varepsilon)} \tau)(\psi_{z'}^n)| &= |(\Pi_{z'}^{(\varepsilon)} \Gamma_{z_z}^{(\varepsilon)} \tau)(\psi_{z'}^n)| \lesssim 2^{-3n/2} \sum_{\alpha} |z - z'|^{|\tau| - \alpha} 2^{-\bar{\gamma}n} \varepsilon^{\alpha - \bar{\gamma} - (\delta/2)} \\ &\lesssim 2^{-3n/2} \sum_{\alpha} (\lambda + 2^{-n})^{|\tau| - \alpha} 2^{-\bar{\gamma}n} \varepsilon^{\alpha - \bar{\gamma} + \delta} \lesssim 2^{-3n/2} 2^{-\bar{\gamma}n} \varepsilon^{|\tau| - \bar{\gamma} + \delta}, \end{aligned} \quad (6.7)$$

where  $\alpha$  runs over all homogeneities less or equal to  $|\tau|$  appearing in  $\mathcal{U}'$  and the last inequality is a consequence of the fact that we only consider  $\lambda$  and  $n$  such that  $\lambda + 2^{-n} \lesssim \varepsilon$ . (The term corresponding to  $\mathbf{1}$ , for which (6.5) does not hold in principle, does not contribute since  $\psi_{z'}^n$  integrates to 0.) The same bound (with  $n$  replaced by  $N$ ) can also be obtained for  $|(\Pi_z^{(\varepsilon)} \tau)(\varphi_{z'}^N)|$ . (In that case  $\varphi_{z'}^N$  does not integrate to 0, but since  $2^{-N} \approx \varepsilon$ , the contribution arising from  $\mathbf{1}$  is of order  $2^{-3N/2} \varepsilon^{|\tau|}$  which is in particular bounded by the right hand side of (6.7).)

It remains to note that, for fixed  $n$ , the number of nonvanishing values of  $A_{z'}^{n, \psi}$  (or  $A_{z'}^N$ ) is of order 1 if  $2^{-n} \geq \lambda$  and of order  $(\lambda/2^{-n})^3$  otherwise. Combining all of these bounds and using the fact that  $\bar{\gamma} \in (0, 1)$ , we finally obtain

$$\begin{aligned} |(\Pi_z^{(\varepsilon)} \tau)(\eta_z^n)| &\lesssim \sum_{\lambda \leq 2^{-n} \leq \varepsilon} \lambda 2^{(1-\bar{\gamma})n} \varepsilon^{|\tau| - \bar{\gamma} - (\delta/2)} \\ &\quad + \sum_{2^{-n} \leq \lambda} 2^{-\bar{\gamma}n} \varepsilon^{|\tau| - \bar{\gamma} - (\delta/2)} \lesssim \lambda^{\bar{\gamma}} \varepsilon^{|\tau| - \bar{\gamma} - (\delta/2)}, \end{aligned}$$

as required. □

As a consequence, we obtain the following Kolmogorov-type convergence criterion.

**PROPOSITION 6.3.** Let  $(\mathcal{T}, \mathcal{G})$  be the regularity structure built in Section 3 and let  $\mathfrak{M}_\varepsilon = (\hat{\Pi}^{(\varepsilon)}, \hat{f}^{(\varepsilon)})$  be as in Theorem 6.1. Assume that there exists  $\delta > 0$  such that, for every test function  $\eta \in \mathcal{B}$ , every  $\tau \in \bar{\mathcal{W}}$  with  $|\tau| < 0$ , every  $x \in \mathbf{R}^2$  and every  $\lambda \in (0, 1]$  there exists a random variable  $(\hat{\Pi}_z^\lambda \tau)(\eta_z^\lambda)$  such that

$$\mathbf{E}|(\hat{\Pi}_z^{(\varepsilon)} \tau)(\eta_z^\lambda)|^2 \lesssim \lambda^{2|\tau|+\delta}, \quad \mathbf{E}|(\hat{\Pi}_z \tau - \hat{\Pi}_z^{(\varepsilon)} \tau)(\eta_z^\lambda)|^2 \lesssim \varepsilon^\delta \lambda^{2|\tau|+\delta}. \quad (6.8a)$$

Assume furthermore that, for  $\tau$  with  $\mathcal{E}^k(\tau) \in \mathcal{W}_+$ , one has

$$\mathbf{E}|D_z^\ell \hat{f}_z^{(\varepsilon)}(\mathcal{E}_0^k(\tau))| \lesssim \varepsilon^{|\tau|+k-|\ell|+\delta}, \quad (6.8b)$$

and that, for  $\tau \in \mathcal{U}'$ , one has the bound

$$\mathbf{E}|(\hat{\Pi}_z^{(\varepsilon)} \tau)(\eta_z^\lambda)| \lesssim \lambda^{\bar{\gamma}+\delta} \varepsilon^{|\tau|-\bar{\gamma}+\delta}, \quad (6.8c)$$

for  $\lambda \leq \varepsilon$  and for test functions  $\eta$  that integrate to 0. Then, there exists a random model  $(\hat{\Pi}, \hat{f}) \in \mathcal{M}_0$  such that  $|\hat{\Pi}^{(\varepsilon)}; \hat{\Pi}|_\varepsilon \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$ .

*Proof.* The proof goes in two steps: first, we show that there is a limiting model  $(\hat{\Pi}, \hat{f})$  such that  $|\hat{\Pi}^{(\varepsilon)}; \hat{\Pi}| \rightarrow 0$  in probability, and then we show that  $\|\hat{\Pi}^{(\varepsilon)}\|_\varepsilon \rightarrow 0$  in probability. If we restrict ourselves to the sector  $\mathcal{T}_- \subset \mathcal{T}$  spanned by basis vectors in  $\mathcal{W}$  with negative (or vanishing) homogeneity, the first step follows in exactly the same way as in [Hai14, HP15], using [Hai14, Theorem 10.7]. This by itself does however not yet yield convergence on all of  $\mathcal{T}$ . The reason for this is that it contains basis vectors of the form  $\bar{\tau} = \mathcal{E}^k(\tau)$  with  $|\mathcal{E}^k(\tau)| > 0$ . These do not satisfy the assumptions of [Hai14, Proposition 3.31] since one does not have any *a priori* control over the components of  $\Gamma_{z,z'} \bar{\tau}$ . (Unlike in [Hai14, HP15] where, for vectors of the form  $\bar{\tau} = \mathcal{I}(\tau)$ , such a control was given by [Hai14, Theorem 5.14].)

Note now that, by the definition (3.11), all of the vectors of the form  $\tau = \mathcal{E}^k(\bar{\tau})$  appearing in  $\bar{\mathcal{W}}$  have  $|\bar{\tau}| < 0$  (or  $\bar{\tau} = \mathbf{1}$ , but this case is trivial). By simple inspection, we see that those vectors such that furthermore  $|\tau| > 0$  are necessarily of the form

$$\tau = \mathcal{E}^{j-1}(\bar{\tau}), \quad \bar{\tau} = \Psi^{2j-n} \mathcal{I}'(\mathcal{E}^{k_1-1} \Psi^{2k_1}) \cdots \mathcal{I}'(\mathcal{E}^{k_\ell-1} \Psi^{2k_\ell}), \quad (6.9)$$

with  $n > 2$ ,  $j \in \{[n/2], \dots, m\}$ , and  $k_i \in \{1, \dots, m\}$ . At this stage, we note that since  $|\mathcal{I}'(\mathcal{E}^{k-1} \Psi^{2k})| < 0$ ,  $|\mathcal{E}^{k-1} \Psi^{2k}| < 0$ , and  $|\Psi| < 0$ , the structure group acts trivially on  $\bar{\tau}$ , so that one has the identity

$$\hat{f}_z^{(\varepsilon)}(\mathcal{E}_\ell^k(\bar{\tau})) = D_z^\ell \hat{f}_z^{(\varepsilon)}(\mathcal{E}_0^k(\bar{\tau})). \quad (6.10)$$

Setting  $g(z) = \hat{f}_z^{(\varepsilon)}(\mathcal{E}_0^k(\bar{\tau}))$  as a shorthand, it then follows from (3.30) that

$$\hat{\gamma}_{z\bar{z}}^{(\varepsilon)}(\mathcal{E}_\ell^k(\bar{\tau})) = g(\bar{z}) - \sum_{|m| < |\tau| + k - |\ell|} \frac{(\bar{z} - z)^m}{m!} D^{(m)} g(z). \quad (6.11)$$

It follows immediately from the Kolmogorov continuity test combined with (6.10) that the bound (6.8b) implies not only that (6.4) holds, but also that the required convergence of  $\hat{\Pi}^{(\varepsilon)}$  holds on every element  $\tau$  of the form (6.9). Through (6.11), it also yields the missing bound on  $\hat{\Gamma}_{z\bar{z}}^{(\varepsilon)}$  on all of  $\mathcal{T}$ . From this point on, the proof that  $|\hat{\Pi}^{(\varepsilon)}; \hat{\Pi}| \rightarrow 0$  in probability as  $\varepsilon \rightarrow 0$  proceeds in exactly the same fashion as the proof of [Hai14, Theorem 10.7].

Since we have furthermore already shown that (6.4) holds, it only remains to show that (6.5) holds as well. This however follows immediately from (6.8c), using the equivalence of moments of random variables belonging to a fixed Wiener chaos in the same way as in [Hai14, Theorem 10.7], combined with the fact that wavelet basis functions do indeed integrate to 0.  $\square$

## 6.2. Proof of Theorem 6.1.

*Proof.* As a consequence of Proposition 6.3, we only need to show that the bounds (6.8) hold. We start with the proof that (6.8a) holds. Actually, as a consequence of [Hai14, Theorem 5.14], we only require these bounds for symbols that are not of the form  $\mathcal{I}(\tau)$  or  $\mathcal{I}'(\tau)$ . Furthermore, it suffices to show (6.8a) for  $z = 0$  by translation invariance, and most of this section is devoted to this proof. We first consider those basis vectors that do not contain the symbol  $\mathcal{E}$ .

**6.2.1. The case  $\tau = \mathbb{V}$ .** The first nontrivial symbol is given by  $\tau = \mathbb{V}$ . In order to represent the random variable  $(\hat{\Pi}_0^{(\varepsilon)}\tau)(\varphi_0^\lambda)$  for some test function  $\varphi$ , we make use of the following graphical notation, which is essentially the same as in [HP15]. Elements belonging to the  $k$ th Wiener chaos are represented by kernels with  $k$  space–time arguments, via the map  $f \mapsto I_k(f)$  described in [Nua06, Ch. 1.1.2]. We will sometimes represent such a kernel by a graph which contains  $k$  distinguished vertices of the type  $\circ$ , each of them representing one of the arguments of the kernel. A special vertex  $\bullet$  represents the origin 0. All other vertices represent integration variables.

Each line then represents a kernel, with  $\longrightarrow$  representing the kernel  $K'$ ,  $\dashrightarrow$  representing the kernel  $K'_\varepsilon = \varrho_\varepsilon * K'$ , and  $\longrightarrow$  representing a generic test function  $\varphi_0^\lambda$  rescaled to scale  $\lambda$ . Whenever we draw a barred arrow  $\longrightarrow$  this represents a factor  $K'(\bar{z} - z) - K'(-z)$ , where  $z$  and  $\bar{z}$  are the coordinates of the starting and end point, respectively.

With these graphical notations at hand, we have the following expression for the unrenormalized model  $\Pi_0^{(\varepsilon)}$ :

$$(\Pi_0^{(\varepsilon)} \mathbf{v})(\varphi_0^\lambda) = \text{Diagram 1} + \text{Diagram 2}.$$

Here, via the correspondence explained above, the first term represents the element  $I_2(f)$  of the second Wiener chaos associated to the kernel

$$f(z_1, z_2) = \int \varphi_0^\lambda(z) K'_\varepsilon(z - z_1) K'_\varepsilon(z - z_2) dz,$$

while the second term represents the constant

$$\iint \varphi_0^\lambda(z) (K'_\varepsilon(z - \bar{z}))^2 d\bar{z} dz.$$

All variables  $z$ ,  $\bar{z}$ , and so on, appearing in these expressions denote space–time variables.

At this stage, we realize that for  $\varepsilon$  sufficiently small, the second term is identical to  $C_0^{(\varepsilon)} \int \varphi_0^\lambda(z) dz = C_0^{(\varepsilon)} (\Pi_0^{(\varepsilon)} \mathbf{1})(\varphi_0^\lambda)$  with  $C_0^{(\varepsilon)}$  given by (6.1). As a consequence, this term cancels out exactly in the definition of  $\hat{\Pi}_0^{(\varepsilon)} \mathbf{v}$  and we have

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbf{v})(\varphi_0^\lambda) = \text{Diagram 3}. \quad (6.12)$$

We now argue that we can find random variables  $(\Pi_0 \mathbf{v})(\varphi_0^\lambda)$  so that the bound (6.8a) does indeed hold. Note first that as a consequence of [Nua06, Ch. 1.1.2] and of the fact that symmetrization is a projection in  $L^2$ , a random variable  $X$  belonging to the  $k$ th homogeneous Wiener chaos and represented by a kernel  $K_X$  satisfies  $\mathbf{E}X^2 \leq k! \|K_X\|_{L^2}^2$ . As a consequence of (6.12), one therefore has the bound

$$\mathbf{E}|(\hat{\Pi}_0^{(\varepsilon)} \mathbf{v})(\varphi_0^\lambda)|^2 \leq 2 \times \text{Diagram 4}. \quad (6.13)$$

Furthermore, using the explicit form of the heat kernel, one can verify that the kernel  $K'_\varepsilon$  satisfies

$$\sup_{\varepsilon \in (0,1]} \|K'_\varepsilon\|_{2;p} < \infty,$$

where  $\|\cdot\|_{\alpha;p}$  is given by (A.1) below. (In particular, it also satisfies the same bound with 2 replaced by  $2 + \kappa/2$ .) The right hand side of (6.13) is therefore precisely of the form  $\mathcal{I}_\lambda^\mathcal{G}(K)$  for some collection of kernels  $K$  satisfying the assumptions of Section A uniformly over  $\varepsilon \in (0, 1]$  and for the labelled graph  $\mathcal{G}$  given by



(Here, the label  $(2+, 0)$  on an edge  $e$  means that  $a_e = 2 + \kappa/2$  and  $r_e = 0$ .) It is straightforward to verify that the assumptions of Theorem A.3 are satisfied, so that one has the bound  $\mathbf{E}[(\hat{\Pi}_0^{(\varepsilon)} \mathbf{v})(\varphi_0^\lambda)^2] \lesssim \lambda^\alpha$  where

$$\alpha = \#\{\text{vertices not adjacent to root}\}|\mathbf{s}| - \sum_e a_e = 2 \cdot 3 - 8 - 2\kappa = 2|\mathbf{v}| + 2\kappa,$$

since the homogeneity of  $\mathbf{v}$  is  $|\mathbf{v}| = -1 - 2\kappa$ . This bound holds uniformly over  $\varepsilon \in (0, 1]$ , so that it is indeed the required first bound in (6.8a).

REMARK 6.4. From now on, whenever we write  $\mathcal{I}_\lambda^\mathcal{G}$  without specifying a collection of kernels  $K$ , we really mean  $\mathcal{I}_\lambda^\mathcal{G}(K)$  for a collection of kernels  $K$  satisfying the assumptions of Section A uniformly over  $\varepsilon \in (0, 1]$ .

We still need to obtain the second bound in (6.8a). This however can be obtained in exactly the same way as soon as we note that, when considering the difference between  $\Pi_0$  and  $\hat{\Pi}_0^{(\varepsilon)}$ , we obtain a sum of expressions of the type (6.12), but in each term some of the instances of  $K'_\varepsilon$  are replaced by  $K'$  and exactly one instance is replaced by  $K'_\varepsilon - K'$ . We then use the fact that  $K'$  satisfies the same bound as  $K'_\varepsilon$ , while  $K'_\varepsilon - K'$  satisfies the improved bound

$$\|K'_\varepsilon - K'\|_{2+\kappa/2;p} \lesssim \varepsilon^{\kappa/2},$$

as a consequence of [Hai14, Lemma 10.17]. This is the reason for using labels  $2 + \kappa/2$  in (6.14) rather than 2, since although  $\sup_{\varepsilon \in (0,1]} \|K'_\varepsilon\|_{2;p} < \infty$ , one has  $\|K'_\varepsilon - K'\|_{2;p} \not\rightarrow 0$  as  $\varepsilon \rightarrow 0$ . This is the same for all of the symbols, so we only ever explicitly show how to obtain the first bound in (6.8a) with the understanding that the second bound then follows in the same way.

6.2.2. *The case  $\tau = \diamond$ .* We now turn to  $\tau = \diamond$ . This time, one has  $\hat{\Pi}_0^{(\varepsilon)} \diamond = \Pi_0^{(\varepsilon)} \diamond$  so that, similarly to before, we have the identity

$$(\hat{\Pi}_0^{(\varepsilon)} \diamond)(\varphi_0^\lambda) = \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} + \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} = \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} - \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} \quad (6.15)$$

In order to see this, recall that the barred arrow represents a difference  $K'(\bar{z} - z) - K'(-z)$ , so that one has the identity

$$\begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} = \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} - \begin{array}{c} \circ \quad \circ \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array}$$

The first term appearing on the right hand side of this expression vanishes because the kernel  $(K' * K'_\varepsilon) \cdot K'$  is odd under the substitution  $(t, x) \mapsto (t, -x)$  (recall that we assumed that the mollifier  $\varrho$  is even under that substitution), so that it integrates to 0, thus yielding (6.15).

Since random variables belonging to Wiener chaoses of different order are orthogonal, we obtain as before the bound

$$\mathbf{E}|(\hat{\Pi}_0^{(\varepsilon)} \diamond)(\varphi_0^\lambda)|^2 \leq 2 \begin{array}{c} \bullet \quad \bullet \quad \bullet \\ \downarrow \quad \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} + \left( \begin{array}{c} \bullet \quad \bullet \\ \downarrow \quad \downarrow \\ \bullet \rightarrow \bullet \\ \uparrow \\ \bullet \end{array} \right)^2.$$

Both terms separately can be bounded in the same way as before. This time however the first term is given by  $\mathcal{I}_\lambda^{\mathcal{G}}$  for the graph

$$\mathcal{G} = \begin{array}{c} \bullet \xleftarrow{2+,0} \bullet \xrightarrow{2+,0} \bullet \\ \downarrow \scriptstyle 2,1 \quad \downarrow \scriptstyle 2,1 \\ \bullet \xleftarrow{2+,0} \bullet \xrightarrow{2+,0} \bullet \\ \searrow \quad \swarrow \\ \bullet \end{array},$$

that is, the two vertical edges have  $r_e = 1$ . Again, it is straightforward to verify that Assumption A.1 is verified, so that the required bounds follow.

6.2.3. *The case  $\tau = \mathbb{V}$ .* We now turn to  $\tau = \mathbb{V}$  which is slightly trickier. One can verify from its recursive definition that the structure group acts trivially on  $\mathbb{V}$ , so that one has similarly to before the identity

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(\varphi_0^\lambda) = \text{[diagram]} + 2 \text{[diagram]}, \quad (6.16)$$

where the second term comes from the product formula [Nua06]. This time, it turns out that when trying to ‘naively’ apply Theorem A.7, its conditions fail to be satisfied for the second term. Denote however by  $Q_\varepsilon$  the kernel

$$Q_\varepsilon(z) = \text{[diagram]} = K(z) \int K_\varepsilon(z - \bar{z}) K_\varepsilon(-\bar{z}) d\bar{z}.$$

It then follows by symmetry as above that  $\int Q_\varepsilon(z) dz = 0$ . As a consequence, for any  $\varepsilon > 0$ , the distribution  $\mathcal{R}Q_\varepsilon(z)$  given by (A.5) with  $I_{e,k} = 0$  is exactly *the same* as simple integration against  $Q_\varepsilon$ , without any renormalization. Furthermore, it follows easily from [Hai14, Section 10] that there is a limiting kernel  $Q$  such that  $\sup_{\varepsilon \in (0,1]} \|Q_\varepsilon\|_{3,p} < \infty$  and  $\|Q_\varepsilon - Q\|_{3+\kappa,p} \lesssim \varepsilon^\kappa$ . Writing  $\text{[diagram]}$  as a graphical notation for the kernel  $Q_\varepsilon = \mathcal{R}Q_\varepsilon$ , we can rewrite (6.16) as

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(\varphi_0^\lambda) = \text{[diagram]} + 2 \text{[diagram]}$$

and bound  $\mathbf{E}[(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(\varphi_0^\lambda)]^2$  by a constant multiple of  $|\mathcal{I}_\lambda^\mathcal{G}| + |\mathcal{I}_\lambda^{\bar{\mathcal{G}}}|$  for graphs  $\mathcal{G}$  and  $\bar{\mathcal{G}}$  given by

$$\mathcal{G} = \text{[diagram]}, \quad \bar{\mathcal{G}} = \text{[diagram]}.$$

Again, Assumption A.1 can easily be checked for both of these graphs so that, in view of the above comments, Theorem A.7 applies and yields the desired bounds.



6.2.4. *The case  $\tau = \mathbb{V}\mathbb{V}$ .* Again, the structure group acts trivially on  $\mathbb{V}\mathbb{V}$  and one has the identity  $\Delta^{\text{Wick}} M_0 \mathbb{V}\mathbb{V} = (\mathbb{V}\mathbb{V} - C_3^{(\varepsilon)} \mathbf{1}) \otimes \mathbf{1}$ . As a consequence, we obtain the identity

$$\hat{\Pi}_0^{(\varepsilon)} \mathbb{V}\mathbb{V} = (K' * \Pi_0^{(\varepsilon)} \mathbb{V})^2 - 2C_3^{(\varepsilon)}.$$

When testing against the test function  $\varphi_0^\lambda$ , it follows from the product formula and the definition of  $C_3^{(\varepsilon)}$  that the Wiener chaos decomposition of this expression is given by

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V}\mathbb{V})(\varphi_0^\lambda) = \text{Diagram 1} + 4 \text{Diagram 2}.$$

Note that the term appearing in the Wiener chaos of order 0 is cancelled out exactly by the renormalization constant  $C_3^{(\varepsilon)}$ , which is why it does not appear here. Similarly to before, it is now straightforward to reduce ourselves to the situation of Theorem A.7 and to verify that Assumption A.1 holds for the two resulting labelled graphs.

6.2.5. *The case  $\tau = \mathbb{V}\mathbb{V}_\tau$ .* This time the structure group acts nontrivially on  $\mathbb{V}\mathbb{V}_\tau$  and it follows from (3.20d) combined with the definition of the renormalization map  $M^{(\varepsilon)}$  that

$$\hat{\Pi}_0^{(\varepsilon)} \mathbb{V}\mathbb{V}_\tau = ((K' * \hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(\cdot) - (K' * \hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(0))(K' * \xi^{(\varepsilon)}) - 2C_2^{(\varepsilon)}.$$

As a consequence, one has the identity

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V}\mathbb{V}_\tau)(\varphi_0^\lambda) = \text{Diagram 1} + \text{Diagram 2} + 2 \text{Diagram 3} + 2 \text{Diagram 4} + 2 \text{Diagram 5} - 2C_2^{(\varepsilon)} \text{Diagram 6}.$$

At this stage we note again that the last two terms cancel each other out, except for the fact that one of the arrows in the penultimate term is ‘barred’. Using again

the notation  $\rightsquigarrow$  for the kernel  $Q_\varepsilon$ , we can therefore rewrite this as

$$(\hat{\Pi}_0^{(\varepsilon)} \mathbb{V})(\varphi_0^\lambda) = \text{diagram 1} + \text{diagram 2} - \text{diagram 3} + 2 \cdot \text{diagram 4} + 2 \cdot \text{diagram 5} - 2 \cdot \text{diagram 6}.$$

At this stage, we can once again reduce ourselves to the situation of Theorem A.7 just as above.

6.2.6. *Symbols containing  $\mathcal{E}$*  We now turn to the proof of (6.8a) for those symbols  $\tau$  with  $|\tau| < 0$  which contain at least one occurrence of the symbol  $\mathcal{E}$ .

We first consider symbols of the type  $\tau = \mathcal{E}^k(\Psi^{2k+2})$ . Note that for  $k = 0$ , one has  $\tau = \mathbb{V}$ , which has already been treated, so we assume that  $k \geq 1$ . Thanks to (5.6), the choice of renormalization constant in the definition of  $M^{(\varepsilon)}$ , and the definition of the Wick product, one has the identity

$$(\hat{\Pi}_0^{(\varepsilon)} \tau)(\varphi_0^\lambda) = (\hat{\Pi}_0^{(\varepsilon)} \mathcal{E}^k(\Psi^{2k+2}))(\varphi_0^\lambda) = \varepsilon^k ((\hat{\Pi}_0^{(\varepsilon)} \Psi)^{\diamond(2k+2)})(\varphi_0^\lambda),$$

which can also be written as

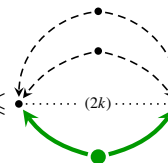
$$(\hat{\Pi}_0^{(\varepsilon)} \tau)(\varphi_0^\lambda) = \varepsilon^k \text{diagram} . \quad (6.17)$$

At this stage, we introduce the shorthand  $\cdots$  for the kernel  $N_\varepsilon \stackrel{\text{def}}{=} \varepsilon(K' * Q_\varepsilon) * (K' * Q_\varepsilon)(-)$ , namely

$$\bullet \cdots \bullet = \mathcal{E} \bullet \leftarrow \cdots \bullet \rightarrow \bullet.$$

(Since this kernel is symmetric, its orientation is irrelevant so we do not draw any arrow on it.) With this notation, we then obtain in a similar way to before

the bound

$$\mathbf{E} |(\hat{\Pi}_0^{(\varepsilon)} \tau)(\varphi_0^\lambda)|^2 \leq \dots (2k) \dots \quad (6.18)$$


where we wrote  $\dots (2k) \dots$  as a shorthand for  $N_\varepsilon^{2k}$  on the right. We also note that the scaling properties of  $K'$  yield the bound

$$\|N_\varepsilon\|_{\delta; p} \lesssim \varepsilon^\delta, \quad (6.19)$$

for every  $\delta \in (0, 1]$  and every  $p > 0$ . Indeed, it suffices to note that [Hai14, Lemma 10.17] implies that  $\|K' * Q_\varepsilon\|_{1+\delta; p} \lesssim \varepsilon^{\delta-1}$  which then allows us to apply [Hai14, Lemma 10.14]. As a consequence, we are again in the setting of Theorem A.7, with a graph  $\tilde{\mathcal{G}}$  that is exactly the same as the graph  $\mathcal{G}$  in (6.14), except for an additional edge with  $a_e$  arbitrarily small connecting the left and right vertices. Since Assumption A.1 is an open condition any graph  $\tilde{\mathcal{G}}$  obtained from another graph  $\mathcal{G}$  by the addition of some new edges with  $a_e = \delta$  or the increase of the homogeneities of some edges by  $\delta$  satisfies Assumption A.1 for  $\delta$  sufficiently small, provided that the original graph  $\mathcal{G}$  satisfies it. Combining this with (6.19), it follows that one has the bound

$$\mathbf{E} |(\hat{\Pi}_0^{(\varepsilon)} \tau)(\varphi_0^\lambda)|^2 \lesssim \varepsilon^\delta \lambda^{|\tau|+\delta},$$

for some sufficiently small choice of  $\delta$ . In particular, the bounds (6.8a) are satisfied with  $\hat{\Pi}_0 \tau = 0$ .

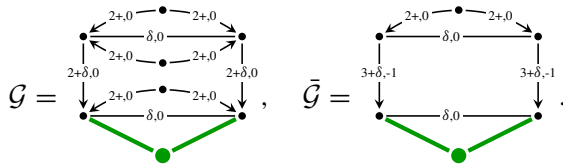
Something similar happens for all other symbols containing at least one instance of  $\mathcal{E}$ . Indeed, consider next  $\tau = \mathcal{E}^k(\Psi^{2k+1} \mathcal{I}'(\mathcal{E}^\ell \Psi^{2\ell+2}))$  with  $k, \ell \geq 0$ , which is the ‘ $\mathcal{E}$ -decorated’ version of  $\tau = \mathbb{V}$ . As a consequence of (5.6), Proposition 5.7, and the fact that  $\hat{\Pi}^{(\varepsilon)}$  is again an admissible model by construction (see [Hai14, Section 8]), we conclude that one has the identity

$$(\hat{\Pi}_0^{(\varepsilon)} \tau)(z) = \varepsilon^{k+\ell} (K' * \xi^{(\varepsilon)})(z) \diamond^{(2k+1)} (K' * (K' * \xi^{(\varepsilon)})^{\diamond(2\ell+2)})(z), \quad (6.20)$$

where we use the symbol  $\diamond$  to denote the Wick product (or rather Wick power in this case), see [Nua06]. Similarly to above, the kernel  $Q_\varepsilon^{(m)} = Q_\varepsilon N_\varepsilon^m$  is odd for every  $m \geq 0$ , so that it can again be identified with  $\mathcal{R} Q_\varepsilon^{(m)}$ . Furthermore, it is of order  $3 + \delta$  for any  $\delta > 0$  and  $\|Q_\varepsilon^{(m)}\|_{3+\delta, p} \lesssim \varepsilon^\delta$  for  $\delta \in (0, 1)$  provided that  $m > 0$ . Combining this with (6.20) we conclude that, for every sufficiently small exponent  $\delta, \bar{\delta} > 0$ , one has again a bound of the type

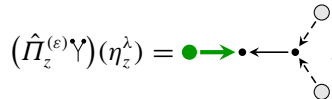
$$\mathbf{E} |(\hat{\Pi}_0^{(\varepsilon)} \tau)(\varphi_0^\lambda)|^2 \leq |\mathcal{I}_\lambda^\mathcal{G}| + |\mathcal{I}_\lambda^{\bar{\mathcal{G}}}|,$$

but this time the two labelled graphs  $\mathcal{G}$  and  $\bar{\mathcal{G}}$  are given by

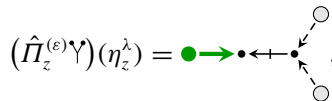


Furthermore, again as a consequence of the bound (6.19) and the corresponding bound for  $Q_\varepsilon^{(m)}$ , it follows that as soon as  $k + \ell > 0$ , at least one of the factors  $\|K_e\|_{a_e; p}$  appearing in Theorem A.7 is bounded by  $\varepsilon^\delta$ , thus yielding the required bound.

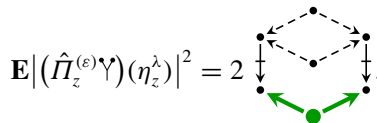
6.2.7. *Additional bounds on  $\hat{\Pi}^{(\varepsilon)}$*  We now turn to the proof of the bound (6.8c). This bound is of course nontrivial only for symbols  $\tau$  with  $|\tau| < \bar{\gamma}$ . The bound for  $\tau = \uparrow$  is very easy to obtain so we do not dwell on it. Regarding  $\tau = \Upsilon$ , we can write it as in (6.12) as



Since the test function  $\eta$  integrates to 0, this is *equal* to



so that we have the bound



At this point, we note that, as a consequence of [Hai14, Lemma 10.7], we have the bound

$$|K'_\varepsilon|_{\alpha; p} \lesssim \varepsilon^{\alpha-2},$$

for every  $\alpha \in [1, 2]$ . For such values of  $\alpha$ , we can therefore write

$$\mathbf{E} |(\hat{\Pi}_z^{(\varepsilon)} \Upsilon)(\eta_z^\lambda)|^2 \lesssim \varepsilon^{4(\alpha-2)} |\mathcal{I}_\lambda^{\mathcal{G}}|, \quad \mathcal{G} = \begin{array}{c} \bullet \\ \swarrow \alpha,0 \quad \searrow \alpha,0 \\ \bullet \quad \bullet \\ \swarrow \alpha,0 \quad \searrow \alpha,0 \\ \bullet \quad \bullet \\ \downarrow 2,1 \quad \downarrow 2,1 \\ \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}$$

One can now verify that as long as  $\alpha > \frac{3}{2}$ , the conditions of Theorem A.7 are satisfied, so that one has the bound

$$\mathbf{E} |(\hat{\Pi}_z^{(\varepsilon)} \Upsilon)(\eta_z^\lambda)|^2 \lesssim \varepsilon^{4\alpha-8} \lambda^{8-4\alpha}.$$

In particular, since  $\bar{\gamma} < 1$ , we can choose  $\alpha$  such that  $8 - 4\alpha = 2\bar{\gamma} + \kappa$ , so that the required bound (6.8c) follows for  $\tau = \Upsilon$ .

We now turn to  $\tau = \check{\Upsilon}$ . Following the exact same procedure, combined with the steps from Section 6.2.3, we see that in this case one has

$$\mathbf{E} |(\hat{\Pi}_z^{(\varepsilon)} \check{\Upsilon})(\eta_z^\lambda)|^2 \lesssim \varepsilon^{2(\alpha-2)} (|\mathcal{I}_\lambda^{\mathcal{G}}| + |\mathcal{I}_\lambda^{\bar{\mathcal{G}}}|),$$

where the graphs  $\mathcal{G}$  and  $\bar{\mathcal{G}}$  are given by

$$\mathcal{G} = \begin{array}{c} \bullet \\ \swarrow 2,0 \quad \searrow 2,0 \\ \bullet \quad \bullet \\ \swarrow 2,0 \quad \searrow 2,0 \\ \bullet \quad \bullet \\ \swarrow \alpha,0 \quad \searrow \alpha,0 \\ \bullet \quad \bullet \\ \downarrow 2,1 \quad \downarrow 2,1 \\ \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}, \quad \bar{\mathcal{G}} = \begin{array}{c} \bullet \\ \swarrow 2,0 \quad \searrow 2,0 \\ \bullet \quad \bullet \\ \swarrow \alpha+1,0 \quad \searrow \alpha+1,0 \\ \bullet \quad \bullet \\ \downarrow 2,1 \quad \downarrow 2,1 \\ \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}.$$

Again, one can verify that the assumptions of Theorem A.7 hold provided that we choose  $\alpha > \frac{3}{2}$  so that we then obtain the bound

$$\mathbf{E} |(\hat{\Pi}_z^{(\varepsilon)} \check{\Upsilon})(\eta_z^\lambda)|^2 \lesssim \varepsilon^{2\alpha-4} \lambda^{5-2\alpha}.$$

Again, the required bound follows since  $\bar{\gamma} < 1$ . The case  $\tau = \check{\check{\Upsilon}}$  follows in a very similar way. All other symbols in  $\mathcal{U}'$  of homogeneity below 1 are just ‘decorated’ versions of  $\mathfrak{t}$ ,  $\Upsilon$ ,  $\check{\Upsilon}$ , or  $\check{\check{\Upsilon}}$  and can therefore be treated in exactly the same way as in Section 6.2.6.

6.2.8. *Bounds on  $\hat{f}^{(\varepsilon)}$*  It now only remains to show that the bounds (6.8b) also hold. For this, we recall from (6.9) that the only symbols  $\tau$  such that  $|\tau| < 0$  and  $|\mathcal{E}^{j-1}(\tau)| > 0$  for some  $j > 1$  are all of the form

$$\tau = \Psi^{2j-n} \mathcal{I}'(\mathcal{E}^{k_1-1} \Psi^{2k_1}) \dots \mathcal{I}'(\mathcal{E}^{k_\ell-1} \Psi^{2k_\ell}), \quad (6.21)$$

with  $n > 2$ ,  $j \in \{[n/2], \dots, m\}$ , and  $k_i \in \{1, \dots, m\}$ . In order to bound  $\hat{f}_z^{(\varepsilon)}(\mathcal{O}_0^{j-1}(\tau))$ , note first that, setting  $\Psi^{(\varepsilon)}(z) = (K' * \xi^{(\varepsilon)})(z)$ , it is a straightforward calculation to show that one has the bounds

$$\mathbf{E}|D^k \Psi^{(\varepsilon)}(z)|^2 \lesssim \varepsilon^{-1-2|k|}, \quad |\mathbf{E} D^k \Psi^{(\varepsilon)}(0) D^k \Psi^{(\varepsilon)}(z)| \lesssim (|z| + \varepsilon)^{-1-2|k|}, \quad (6.22)$$

for every multiindex  $k$ . Let now  $\{k_1, \dots, k_r\}$  be a finite collection of such multiindices and set

$$\Psi_{k_1, \dots, k_r}^{(\varepsilon)}(z) = D^{k_1} \Psi^{(\varepsilon)}(z) \diamond \dots \diamond D^{k_r} \Psi^{(\varepsilon)}(z).$$

Combining this with (6.22) and Lemma 6.9 below, it is not difficult to see that

$$\mathbf{E}|K' * \Psi_{k_1, \dots, k_r}^{(\varepsilon)}(z)|^2 \lesssim \varepsilon^{2-r-2 \sum_{i=1}^r |k_i|}.$$

In particular, setting  $\Phi_\ell^{(\varepsilon)}(z) = (K' * (\Psi^{(\varepsilon)})^{\diamond \ell})(z)$ , one has the bound

$$\mathbf{E}|D^k \Phi_\ell^{(\varepsilon)}(z)|^2 \lesssim \varepsilon^{2-\ell-2|k|}. \quad (6.23)$$

We now note that, for  $\tau$  as in (6.21), one has

$$\hat{f}_z^{(\varepsilon)}(\mathcal{O}_0^{j-1}(\tau)) = \varepsilon^{j-1-\ell+\sum_{i=1}^\ell k_i} \Psi^{(\varepsilon)}(z)^{\diamond(2j-n)} \Phi_{2k_1}^{(\varepsilon)}(z) \dots \Phi_{2k_\ell}^{(\varepsilon)}(z).$$

Combining (6.23) and (6.22) with the generalized Leibniz rule and the equivalence of moments for random variables belonging to a Wiener chaos of finite order, we conclude that

$$\mathbf{E}|D^r \hat{f}_z^{(\varepsilon)}(\mathcal{O}_0^{j-1}(\tau))| \lesssim \varepsilon^{(n/2)-1-|r|}.$$

The bound (6.8b) now follows immediately.  $\square$

**6.3. Behaviour of the renormalization constants.** The goal of this section is to provide precise asymptotic results on the behaviour of the renormalization constants  $C_\tau^{(\varepsilon)}$  for  $\tau \in \mathcal{B}$  appearing in the construction of our model. We have the following convergence result.

THEOREM 6.5. Let  $C_2^{(\varepsilon)}$  and  $C_3^{(\varepsilon)}$  be as in (6.1) and let  $C_\tau^{(\varepsilon)}$  be as in (6.2). Then, there exists a constant  $c \in \mathbf{R}$  depending both on the choice of  $K$  and of the mollifier  $\varrho$  such that

$$\lim_{\varepsilon \rightarrow 0} (C_3^{(\varepsilon)} + 4C_2^{(\varepsilon)}) = c. \quad (6.24)$$

Furthermore, for every  $\tau \in \mathcal{B} \setminus \{\mathbb{V}, \mathbb{V}_\tau\}$ , there exists a constant  $c_\tau \in \mathbf{R}$  such that  $\lim_{\varepsilon \rightarrow 0} C_\tau^{(\varepsilon)} = c_\tau$ , and these constants are independent of the choice of kernel  $K$ .

REMARK 6.6. The statement (6.24) is nontrivial since in general both of these constants diverge logarithmically as  $\varepsilon \rightarrow 0$ , see [Hai13]. Note furthermore that although it is very similar, this theorem does *not* follow immediately from [Hai13, Lemma 6.5] because here we consider space–time regularizations of the noise.

For the remainder of this section, it turns out to be more convenient to work with the rescaled kernel

$$K_{\varepsilon, \varrho}(z) \stackrel{\text{def}}{=} (\varrho * \mathcal{S}_\varepsilon^{(1)} K)(z),$$

where the scaling operator  $\mathcal{S}_\varepsilon^{(\alpha)}$  is defined by

$$(\mathcal{S}_\varepsilon^{(\alpha)} K)(t, x) = \varepsilon^\alpha K(\varepsilon^2 t, \varepsilon x).$$

This is because in the rescaled variables, our kernels will turn out to converge to nontrivial limits, which is something that would not be easily seen in the original variables. Similarly to before,  $K'_{\varepsilon, \varrho}$  denotes the spatial derivative of  $K_{\varepsilon, \varrho}$ .

REMARK 6.7. A simple change of variables shows that (6.1) is still valid if we interpret  $\dashrightarrow$  as an instance of the rescaled kernel  $K'_{\varepsilon, \varrho}$  instead of the kernel  $\varrho_\varepsilon * K'$  and  $\longrightarrow$  as an instance of  $(\mathcal{S}_\varepsilon^{(1)} K)' = \mathcal{S}_\varepsilon^{(2)}(K')$  instead of  $K'$ . We make use of these interpretations for the remainder of this section.

Before we turn to the proof of Theorem 6.5, we provide a number of useful technical results. In order to state our first result, we introduce the family of norms

$$\|F\|_{\alpha, \beta} = \sup_{|z| \leq 1} |z|^\alpha |F(z)| + \sup_{|z| \geq 1} |z|^\beta |F(z)|,$$

and we denote by  $\mathcal{B}_{\alpha, \beta}$  the Banach space consisting of the functions  $F: \mathbf{R}^{d+1} \rightarrow \mathbf{R}$  such that  $\|F\|_{\alpha, \beta} < \infty$ . Here, for  $z = (t, x)$ , we denoted by  $|z| = |x| + \sqrt{|t|}$  its parabolic norm.

REMARK 6.8. It is straightforward to show that  $K_{\varepsilon, \varrho}$  and  $K'_{\varepsilon, \varrho}$  belong to  $\mathcal{B}_{1,0}$  and  $\mathcal{B}_{2,0}$ , respectively, and that, for every  $\kappa > 0$ , they converge to limits in  $\mathcal{B}_{1-\kappa,0}$  and  $\mathcal{B}_{2-\kappa,0}$ , respectively. These limits are given by  $P_\varrho$  and  $P'_\varrho$ , respectively, where  $P_\varrho = P * \varrho$  and  $P$  is the heat kernel without truncation.

Our first preparatory result shows how convolution acts in these spaces.

LEMMA 6.9. Suppose that for  $j = 1, 2$ ,  $F_j$  are functions on  $\mathbf{R}^{d+1}$  with parabolic scaling such that  $F_i \in \mathcal{B}_{\alpha_i, \beta_i}$  with  $\alpha_i < d + 2$ ,  $i = 1, 2$  and  $\beta_1 + \beta_2 > d + 2$ . Then there exists  $C > 0$  such that

$$\|F_1 * F_2\|_{\alpha, \beta} \leq C \|F_1\|_{\alpha_1, \beta_1} \|F_2\|_{\alpha_2, \beta_2}, \quad (6.25)$$

with  $\alpha = 0 \vee (\alpha_1 + \alpha_2 - d - 2)$  and  $\beta = (\beta_1 + \beta_2 - d - 2) \wedge \beta_1 \wedge \beta_2$ .

*Proof.* The condition  $\alpha_i < d + 2$ ,  $i = 1, 2$  is required or the integral defining  $F_1 * F_2$  diverges at small scales. Similarly, we need  $\beta_1 + \beta_2 > d + 2$  for the integral to converge at large scales. By bilinearity, we can (and will from now on) assume that  $\|F_j\|_{\alpha_j, \beta_j} = 1$  for  $j \in \{1, 2\}$ .

Let first  $|z| \leq 1$  and write

$$(F_1 * F_2)(z) = \int_{\mathbf{R}^{d+1}} F_1(y) F_2(z - y) dy. \quad (6.26)$$

We now break the domain of integration into four regions  $\{A_i\}_{i=1}^4$  and we bound it separately in each of them. We set

$$\begin{aligned} A_1 &= \{y : |y| \leq 2|z| \text{ \& } |y| \leq |z - y|\}, \\ A_2 &= \{y : |y| \leq 2|z| \text{ \& } |y| > |z - y|\}, \\ A_3 &= \{y : |y| \in (2|z|, 2)\}, \\ A_4 &= \{y : |y| > 2\}. \end{aligned}$$

For  $y \in A_1$ , since  $|z| \leq |y| + |z - y|$ , we have  $|z - y| \geq |z|/2$ , so that  $|F_1(y)F_2(z - y)| \leq |z|^{-\alpha_2}|y|^{-\alpha_1}$ . Integrating this bound over  $\{|y| \leq 2|z|\}$  yields a bound proportional to  $|z|^{d+2-\alpha_1-\alpha_2}$ . Exchanging the roles of  $y$  and  $z - y$ , we obtain the same bound for the integral over  $A_2$ . For  $y \in A_3$ , have  $|z - y| \geq |y| - |z| \geq |y|/2$  and  $|z - y| \leq 3$ , so that  $|F_1(y)F_2(z - y)| \lesssim |y|^{-\alpha_1-\alpha_2}$ . Integrating this bound over  $A_3$  yields this time a bound proportional to  $1 + |z|^{d+2-\alpha_1-\alpha_2}$ . Finally, on  $A_4$ , we also have  $|z - y| \geq |y|/2$ , but we additionally have  $|y| \geq 2$ , so that this time  $|F_1(y)F_2(z - y)| \lesssim |y|^{-\beta_1-\beta_2}$ . Since  $\beta_1 + \beta_2 > d + 2$  by assumption, this is integrable over  $|y| \geq 2$ , so that we obtain a bound proportional to 1, thus completing the required bound on  $|(F_1 * F_2)(z)|$  for  $|z| \leq 1$ .



For  $|z| \geq 1$ , we break the domain of integration for (6.26) into five regions  $\{B_i\}_{i=1}^5$ , namely

$$\begin{aligned} B_1 &= \{y : |y| \leq 1/2\}, \\ B_2 &= \{y : |z - y| \leq 1/2\}, \\ B_3 &= \{y : |y| \leq 2|z| \text{ \& } |y| \leq |z - y|\} \setminus B_1, \\ B_4 &= \{y : |y| \leq 2|z| \text{ \& } |y| > |z - y|\} \setminus B_2, \\ B_5 &= \{y : |y| > 2|z|\}. \end{aligned}$$

On  $B_1$ , we have  $|z - y| \geq |z| - |y| \geq |z|/2$  so that, since furthermore  $|z| \geq 1$ , one has

$$|F_1(y)F_2(z - y)| \lesssim |z|^{-\beta_2}|y|^{-\alpha_1}. \quad (6.27)$$

Integrating this over  $B_1$  yields a bound of the order  $|z|^{-\beta_2}$  since we assumed that  $\alpha_1 < d + 2$ . In the case of  $B_2$ , we similarly obtain a bound of the order  $|z|^{-\beta_1}$ . On  $B_3$ , we have instead  $|F_1(y)F_2(z - y)| \lesssim |z|^{-\beta_2}|y|^{-\beta_1}$ , which we integrate over  $|y| \in (1/2, 2|z|]$ , so that we obtain a bound of the order of  $|z|^{-\beta_2}(1 + |z|^{d+2-\beta_1})$ . In the same way, the integral over  $B_4$  yields a bound of the order of  $|z|^{-\beta_1}(1 + |z|^{d+2-\beta_2})$ . Finally, for  $y \in B_5$ , we have  $|z - y| \geq |y| - |z| \geq |y|/2$  so that  $|F_1(y)F_2(z - y)| \lesssim |y|^{-\beta_1-\beta_2}$ , thus yielding a bound of the order  $|z|^{d+2-\beta_1-\beta_2}$ . Collecting all of these bounds completes the proof.  $\square$

We also need a slightly stronger conclusion in a special case. In order to formulate this, we introduce the family of norms

$$\|F\|_{\alpha,\beta;1} = \sup_{|z| \leq 1} |z|^\alpha |F(z)| + \sup_{|z| \geq 1} |z|^\beta (|F(z)| + |z| |\nabla_x F(z)| + |z|^2 |\partial_t F(z)|),$$

and we denote by  $\mathcal{B}_{\alpha,\beta;1}$  the Banach space consisting of the functions  $F : \mathbf{R}^{d+1} \rightarrow \mathbf{R}$  such that  $\|F\|_{\alpha,\beta;1} < \infty$ .

**LEMMA 6.10.** *Let  $F_j$  as in Lemma 6.9, but with  $\beta_1 > d + 2 > \beta_2 > 0$ ,  $\alpha_i < d + 2$ , and such that additionally  $\int F_1(z) dz = 0$  and  $\|F_2\|_{\alpha_1,\beta_1;1} < \infty$ . Then, one has the stronger conclusion  $\beta = (\beta_1 + \beta_2 - d - 2) \wedge (\beta_2 + 1)$ .*

*Proof.* We only need to consider  $|z| \geq 2$  say and, as before we want to estimate the integral

$$(F_1 * F_2)(z) = \int_{\mathbf{R}^{d+1}} F_1(y) (F_2(z - y) - F_2(z)) dy. \quad (6.28)$$

The reason why this identity holds is of course that we assumed that  $F_1$  integrates to 0. This time, we break the integral into three regions.

First, we consider the case  $|y| \leq |z|/2$ . In this case, as a simple consequence of our bounds on the derivatives of  $F_2$ , one has

$$|F_2(z - y) - F_2(z)| \lesssim |y| |z|^{-\beta_2-1}.$$

On the other hand, one has

$$\int_{|y| \leq |z|} |y| |F_1(y)| dy \lesssim |z|^{0 \vee (d+3-\beta_1)}.$$

Combining the two yields a bound of the required form. For the integral over the region  $|y| \geq 2|z|$ , we use the ‘brutal’ bound

$$|F_2(z - y) - F_2(z)| \lesssim |z|^{-\beta_2},$$

so that this integral is bounded by  $|z|^{-\beta_2} \int_{|y| \geq |z|} |F_1(y)| dy$ . Since we assumed that  $\beta_1 > d + 2$ , this integral converges and is of order  $|z|^{d+2-\beta_1}$  thus yielding the required bound. Finally, in the region  $|z|/2 \leq |y| \leq 2|z|$ , we bound  $|F_1(y)|$  by  $|z|^{-\beta_1}$ . Since  $\beta_2 < d + 2$ , the integral of  $|F_2(z - y) - F_2(z)|$  over that region can be bounded by  $|z|^{d+2-\beta_2}$ , thus again yielding the correct bound.  $\square$

REMARK 6.11. Lemmas 6.9 and 6.10 immediately extend to the case of arbitrary scalings by replacing each instance of  $d + 2$  by the scaling dimension of the underlying space.

Before we turn to the proof of Theorem 6.5, we define a kernel  $Q_\varepsilon$  by

$$Q_\varepsilon(z) = \begin{array}{c} \bullet \\ \nearrow \quad \searrow \\ \circ \end{array} = \int K'_{\varepsilon,Q}(z - \bar{z}) K'_{\varepsilon,Q}(-\bar{z}) d\bar{z}. \quad (6.29)$$

(Recall Remark 6.7 regarding these notations.) We then have the following result.

LEMMA 6.12. *With  $Q_\varepsilon$  as above, there exist kernels  $R_\varepsilon^{(1)}$ ,  $R_\varepsilon^{(2)}$  and  $\tilde{R}_\varepsilon$  such that*

$$\begin{aligned} 2Q_\varepsilon(z) &= K_{\varepsilon,Q}(z) + K_{\varepsilon,Q}(-z) + R_\varepsilon^{(1)}(z) + (\mathcal{S}_\varepsilon^{(1)} R_\varepsilon^{(2)})(z), \\ \mathcal{S}_\varepsilon^{(2)} K' &= K'_{\varepsilon,Q} + \tilde{R}_\varepsilon, \end{aligned}$$

and such that the bounds

$$\|R_\varepsilon^{(1)}\|_{0,2} + \|R_\varepsilon^{(2)}\|_{0,4} + \|\tilde{R}_\varepsilon\|_{2,3} \leq C,$$

hold for some  $C$  independent of  $\varepsilon \in (0, 1]$ . Furthermore, for every  $\kappa > 0$ , these kernels converge in  $\mathcal{B}_{0,2-\kappa}$ ,  $\mathcal{B}_{0,4}$  and  $\mathcal{B}_{2,3-\kappa}$ , respectively, as  $\varepsilon \rightarrow 0$ . In the case of  $R_\varepsilon^{(1)}$ , the limit is 0 and in the case of  $\tilde{R}_\varepsilon$  it is independent of the choice of  $K$ .

*Proof.* Since  $K'_{\varepsilon,\varrho} = \varrho * S_\varepsilon^{(2)} K'$ , the claim for  $\tilde{R}_\varepsilon$  follows from the general fact that, for any kernel  $\hat{K}$  with  $\|\hat{K}\|_{\alpha,\beta} < \infty$  and  $\alpha < d + 2$ , one has

$$\|\hat{K} - \varrho * \hat{K}\|_{\alpha,\beta+1} \lesssim \|\hat{K}\|_{\alpha,\beta}.$$

Regarding  $R_\varepsilon$ , an explicit calculation shows that if we denote by  $P$  the heat kernel, one has the identity

$$2 \int P'(z - \bar{z}) P'(-\bar{z}) d\bar{z} = P(z) + P(-z).$$

(Simply consider the Fourier transforms of both sides.) Since  $K$  is compactly supported and agrees with  $P$  in some neighbourhood of the origin, this immediately implies that there exists a smooth compactly supported function  $R$  such that

$$2 \int K'(z - \bar{z}) K'(-\bar{z}) d\bar{z} = K(z) + K(-z) + R(z).$$

Rescaling and convolving with  $\varrho^{(2)}$ , we conclude that

$$2Q_\varepsilon(z) = (\varrho^{(2)} * S_\varepsilon^{(1)} K)(z) + (\varrho^{(2)} * S_\varepsilon^{(1)} K)(-z) + S_\varepsilon^{(1)} (\varrho_\varepsilon^{(2)} * R)(z),$$

so that we can set

$$\begin{aligned} R_\varepsilon^{(2)} &= \varrho_\varepsilon^{(2)} * R, \\ R_\varepsilon^{(1)}(z) &= ((\varrho^{(2)} - \varrho) * S_\varepsilon^{(1)} K)(z) + ((\varrho^{(2)} - \varrho) * S_\varepsilon^{(1)} K)(-z). \end{aligned}$$

The required bounds then follow easily. □

LEMMA 6.13. Let  $\tilde{C}_2^{(\varepsilon)}$  and  $\tilde{C}_3^{(\varepsilon)}$  be defined by the identities

$$C_2^{(\varepsilon)} = \text{[Diagram: A path of 4 vertices with a green dot at the bottom vertex]} + \tilde{C}_2^{(\varepsilon)}, \quad C_3^{(\varepsilon)} = \text{[Diagram: A cycle of 4 vertices with a green dot at the bottom vertex]} + \tilde{C}_3^{(\varepsilon)}. \quad (6.30)$$

Then both  $\tilde{C}_2^{(\varepsilon)}$  and  $\tilde{C}_3^{(\varepsilon)}$  converge to finite limits as  $\varepsilon \rightarrow 0$ , and these limits do not depend on the choice of the cutoff kernel  $K$ .

*Proof.* Comparing (6.30) to (6.1) and writing  $\longrightarrow$  for the kernel  $D_\varepsilon \stackrel{\text{def}}{=} \mathcal{S}_\varepsilon^{(2)} K' - K'_{\varepsilon, Q}$ , we have

$$\tilde{C}_2^{(\varepsilon)} = \text{diagram 1} + \text{diagram 2}.$$

At this point, we note that  $K'_{\varepsilon, Q}$ ,  $\mathcal{S}_\varepsilon^{(2)} K'$ , and  $D_\varepsilon$  converge in  $\mathcal{B}_{0, 2-\kappa}$ ,  $\mathcal{B}_{2+\kappa, 2-\kappa}$ , and  $\mathcal{B}_{2+\kappa, 3-\kappa}$ , respectively, and that these limits do not depend on the choice of cutoff  $K$ . The claim for  $\tilde{C}_2^{(\varepsilon)}$  now follows by repeatedly applying Lemma 6.9. The constant  $\tilde{C}_3^{(\varepsilon)}$  can be dealt with in a very similar fashion.  $\square$

We now have finally all the ingredients required for the proof of Theorem 6.5.

*Proof of Theorem 6.5.* We first prove that (6.24) holds. Since we also need the kernel  $K_{\varepsilon, Q}$  in this proof, we use for it the graphical notation  $\cdots \longrightarrow$ . As a consequence of Lemmas 6.13, 6.12, and 6.9, we have the identities

$$\begin{aligned} 4C_2^{(\varepsilon)} &= \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \text{diagram 4} + (\dots), \\ 4C_3^{(\varepsilon)} &= 2 \cdot \text{diagram 5} + 2 \cdot \text{diagram 6} + (\dots), \end{aligned} \quad (6.31)$$

where  $(\dots)$  denotes an expression that converges to a finite limit as  $\varepsilon \rightarrow 0$ . This can easily be shown in a way similar to the proof of Lemma 6.13. For example, one of the additional terms appearing in the right hand side of  $C_2^{(\varepsilon)}$  is given by

$$((R_\varepsilon^{(1)} \cdot K'_{\varepsilon, Q}) * K'_{\varepsilon, Q} * P_\varepsilon)(0) + ((R_\varepsilon^{(2)} \cdot K'_\varepsilon) * K'_\varepsilon * K'_\varepsilon * K'_\varepsilon(\cdots))(0). \quad (6.32)$$

To show that this converges to a finite limit, one uses the fact that, by Remark 6.8 and Lemma 6.12,  $R_\varepsilon^{(1)} \cdot K'_{\varepsilon, Q}$ ,  $K'_{\varepsilon, Q}$  and  $P_\varepsilon$  converge as  $\varepsilon \rightarrow 0$  in  $\mathcal{B}_{0, 4-\kappa}$ ,  $\mathcal{B}_{0, 2-\kappa}$ , and  $\mathcal{B}_{0, 1-\kappa}$ , respectively, for every  $\kappa > 0$ . It then suffices to take  $\kappa$  sufficiently small and to apply Lemma 6.9 twice to show that the first term in (6.32) converges to a finite limit. Regarding the second term of (6.32), both  $R_\varepsilon^{(2)} \cdot K'_\varepsilon$  and  $K'_\varepsilon$  converge to limits in  $\mathcal{B}_{2+\kappa, 3}$  for any  $\kappa > 0$  so that its convergence can again be reduced to repeated applications of Lemma 6.9. The other terms appearing in the remainder terms of (6.31) can be dealt with in an analogous way.

At this stage, we perform an integration by parts for the integration variable represented by the top-left vertex in the first term for  $C_3^{(\varepsilon)}$ . This yields the exact identity

$$\text{Diagram 1} = -2 \text{Diagram 2},$$

where the factor 2 comes from the fact that the derivative of  $(K_{\varepsilon, \varrho})^2$  (the two arrows linking the two top vertices) equals  $2K_{\varepsilon, \varrho} K'_{\varepsilon, \varrho}$ . Inserting this into the above expression for  $C_3^{(\varepsilon)}$  yields

$$C_3^{(\varepsilon)} = - \text{Diagram 1} + \frac{1}{2} \text{Diagram 2} + (\dots).$$

We now note that the first term in this expression is identical to the first term appearing in the expression for  $4C_2^{(\varepsilon)}$ . As a consequence, we have

$$C_3^{(\varepsilon)} + 4C_2^{(\varepsilon)} = \text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3} + \frac{1}{2} \text{Diagram 4} + (\dots). \quad (6.33)$$

It is therefore sufficient to show that the four terms appearing on the right hand side of this expression all converge to finite limits as  $\varepsilon \rightarrow 0$ .

To bound the first two terms, we use the easily shown fact that the kernel  $K'_{\varepsilon, \varrho}(z)K_{\varepsilon, \varrho}(-z)$  converges to  $P'_\varrho(z)P_\varrho(-z)$  (where we set  $P_\varrho = P * \varrho$ )  $\mathcal{B}_{0, \beta}$  for every  $\beta > 0$ . The fact that these terms converge to finite limits independent of the choice of  $K$  then immediately follows by applying Lemma 6.9 twice. A virtually identical argument allows to deal with the fourth term. Concerning the third term appearing in the right hand side of (6.33), we note that, by Remark 6.8 and Lemma 6.9, the kernel  $F_\varepsilon \stackrel{\text{def}}{=} (K_{\varepsilon, \varrho} K'_{\varepsilon, \varrho}) * K'_{\varepsilon, \varrho}$  converges to a limit in  $\mathcal{B}_{0, 2-\kappa}$  for any  $\kappa > 0$ , and is supported in  $\{(t, x) : t > -C\}$ , for some fixed constant  $C > 0$ . Since the kernel  $K_{\varepsilon, \varrho}$  also has the same support property and converges in  $\mathcal{B}_{0, 1-\kappa}$ , the product  $F_\varepsilon(z)K_{\varepsilon, \varrho}(-z)$  converges in  $\mathcal{B}_{0, 3-\kappa}$  and is supported in  $\{(t, x) : |t| \leq C\}$ . It is straightforward to conclude that such a function is absolutely integrable for  $\kappa$  small enough, and the claim then follows.

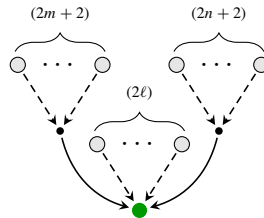
It remains to show that the constants  $C_\tau^{(\varepsilon)}$  have finite limits for all  $\tau \in \mathcal{B} \setminus \{\nabla \nabla, \nabla \cdot\}$ , where  $\mathcal{B}$  was defined in (5.2). Let us first consider elements  $\tau$  of the form

$$\tau = \mathcal{E}^\ell(\Psi^{2\ell} \mathcal{I}'(\mathcal{E}^m(\Psi^{2m+2})) \mathcal{I}'(\mathcal{E}^n(\Psi^{2n+2}))),$$

with  $\ell + m + n > 0$ , which is essentially a ‘decorated’ version of  $\forall\forall$ . By the definition (6.2) of  $C_\tau^{(\varepsilon)}$  combined with the definitions of  $M^{\text{Wick}}$  and  $\Pi^{(\varepsilon)}$ , we have the identity

$$C_\tau^{(\varepsilon)} = \varepsilon^{\ell+m+n} \mathbf{E}((\Psi^{(\varepsilon)})^{\diamond(2\ell)} \Phi_{2m+2}^{(\varepsilon)} \Phi_{2n+2}^{(\varepsilon)})(0),$$

where we used the notations  $\Psi^{(\varepsilon)} = K' * \xi^{(\varepsilon)}$  and  $\Phi_\ell^{(\varepsilon)} = (K' * (\Psi^{(\varepsilon)})^{\diamond\ell})$  as in (6.22) and (6.23). Using graphical notations similar to before and the properties of the Wick product, the expectation appearing in this expression is given by all possible ways of performing pairwise contractions of all nodes of the type  $\circ$  without ever contracting two nodes belonging to the same ‘group’ in the following graph:



It is clear that such a pairing can exist only when no such group is larger than the two others combined, that is, when  $m \leq \ell + n$ ,  $n \leq \ell + m$ , and  $\ell \leq m + n + 2$ . If one of these conditions fails, one has  $C_\tau^{(\varepsilon)} = 0$  and the statement is trivial. If they are satisfied on the other hand, one obtains with the same graphical notations as in (6.18) the identity

$$C_\tau^{(\varepsilon)} = \varepsilon^{\ell+m+n} C_{\ell,m,n} \text{ (graphical diagram) }, \quad (6.34)$$

where the integer values  $a$ ,  $b$  and  $c$  are related to  $\ell$ ,  $m$  and  $n$  by  $a + b = 2\ell$ ,  $a + c = 2m + 2$ ,  $b + c = 2n + 2$ , and the combinatorial factor  $C_{\ell,m,n}$  is given by

$$C_{\ell,m,n} = \frac{(2m+2)!(2n+2)!(2\ell)!}{a!b!c!}.$$

The above conditions on  $\ell$ ,  $m$ ,  $n$  precisely guarantee that  $a$ ,  $b$  and  $c$  are positive. In order to show that  $C_\tau^{(\varepsilon)}$  converges to a limit as  $\varepsilon \rightarrow 0$ , we note first that as



the ‘decorated’ version of  $\checkmark$ . In this case, an argument virtually identical to above shows that one has

$$C_{\tau}^{(\varepsilon)} = C_{\ell,m,n} \begin{array}{c} \text{---} (c) \text{---} \\ \text{---} (b) \text{---} \\ \text{---} (a) \text{---} \end{array}, \quad (6.36)$$

but this time the constants  $a, b, c$  satisfy

$$a + c = 2\ell + 1, \quad a + b = 2m + 1, \quad b + c = 2n + 2. \quad (6.37)$$

As before one has  $C_{\tau}^{(\varepsilon)} = 0$  when  $c = 0$  so that we can assume  $c > 0$ . As before, we then have

$$C_{\tau}^{(\varepsilon)} = C_{\ell,m,n} (\tilde{P}_{\varepsilon}^{(a,b)} * P_{\varepsilon}^c)(0),$$

this time with  $\tilde{P}_{\varepsilon}^{(a,b)} = \tilde{P}_{\varepsilon}^{(a)} * \tilde{P}_{\varepsilon}^{(b)}$  which, as before, converges to a limit  $\tilde{P}^{(a,b)}$  in  $\mathcal{B}_{1,\beta}$  for every  $\beta < (1 + a + b) \wedge (2 + a) \wedge (2 + b)$ . The case  $a = b = 0$  is impossible since one has  $a + b \geq 1$ , so assume first  $a > b = 0$ . As before, this implies that  $c \geq 2$ , so that this case is covered by Lemma 6.9 as above. The case where  $a, b, c > 0$  is also covered in exactly the same way as above. This time however, the case  $b > a = 0$  is not the same as the case  $a > b = 0$  since the conditions (6.37) are no longer symmetric under  $a \leftrightarrow b$ . If  $c \geq 2$ , then this case is covered in the same way as before.

However, it can happen this time that  $a = 0$  and  $c = 1$ , which is not covered by Lemma 6.9 anymore. Our assumptions then imply that  $b \geq 3$ , so that  $\tilde{P}_{\varepsilon}^{(b)}$  is integrable. We furthermore exploit the fact that  $\tilde{P}_{\varepsilon}^{(b)}$  is odd, so that it actually integrates to 0 and we are in the setting of Lemma 6.10 with  $\alpha_1 = \alpha_2 = 2$ ,  $\beta_1 = 5 - \kappa$ , and  $\beta_2 = 2 - \kappa$ . This shows that in this case  $\tilde{P}_{\varepsilon}^{(a,b)}$  converges to  $\tilde{P}^{(a)}$  not only in  $\mathcal{B}_{1,\beta}$  for  $\beta < 2$ , but also for all  $\beta < 3$ . Lemma 6.9 now applies to show that the convolution with  $P_{\varepsilon}$  converges in  $\mathcal{B}_{0,0}$ , thus yielding the required convergence and concluding the proof.  $\square$

## 7. Main convergence results

We are now ready to collect the various results from the previous sections in order to prove the main convergence results of this article.

**7.1. Weak asymmetry regime.** We have the following result, which allows us to identify solutions driven by the model  $\mathfrak{M}$  with the Hopf–Cole solutions to the KPZ equation.



PROPOSITION 7.1. Let  $\gamma, \eta$  be as in Theorem 4.16 and let  $H \in \mathcal{D}^{\gamma, \eta}$  be the solution to (4.20) given by Theorem 4.16 for the model  $\mathfrak{M}$  given by Theorem 6.1, and with initial condition  $h_0 \in \mathcal{C}^\eta$ . Then, there exists a constant  $c$  depending only on the choice of cutoff kernel  $K$  such that the function  $h(t, x) = (\mathcal{R}H)(t, x) - \lambda^3 ct$  is almost surely equal to  $h_{\text{HC}}^{(\lambda)}$  with  $\lambda = \hat{a}_1$ .

In order to prove this result, we give an alternative construction of the model  $\mathfrak{M}$ . This will allow us to obtain Proposition 7.1 as an essentially immediate consequence of [HP15, Theorem 4.7]. To formulate this preliminary result, we define  $\tilde{M}^{(\varepsilon)}$  exactly as  $M^{(\varepsilon)}$ , but this time with  $C_\tau = 0$  for every  $\tau$  of the form (5.2) with  $\ell + m + n > 0$ . Using the same notations as above, we then have the following result:

PROPOSITION 7.2. Let  $\xi^{(\varepsilon)}$  be given by (1.11) and consider the sequence of models on  $\mathcal{T}$  given by

$$\tilde{\mathfrak{M}}_\varepsilon = \tilde{M}^{(\varepsilon)} \mathcal{L}_0(\xi^{(\varepsilon)}),$$

with  $\mathcal{L}_0$  defined in Section 3.6. Then, one has  $\tilde{\mathfrak{M}}_\varepsilon \rightarrow \mathfrak{M}$  in  $\mathcal{M}_0$  in probability, where  $\mathfrak{M}$  is the same (random) model as in Theorem 6.1.

REMARK 7.3. Note that in the statement of Proposition 7.2, we consider the lift  $\mathcal{L}_0$  instead of the lift  $\mathcal{L}_\varepsilon$ . Since we furthermore set  $C_\tau = 0$  for every formal expression  $\tau$  containing the symbol  $\mathcal{E}$ , the model  $\tilde{\mathfrak{M}}_\varepsilon$  yields 0 when applied to any formal expression that includes a power of  $\mathcal{E}$ .

*Proof.* By the combined definitions of  $\mathcal{L}_0$  and  $\tilde{M}^{(\varepsilon)}$  (in particular the fact that  $C_\tau = 0$  for every  $\tau$  of the form (5.2) with  $\ell + m + n > 0$ ), the model  $\tilde{\mathfrak{M}}_\varepsilon = (\tilde{\Pi}^{(\varepsilon)}, \tilde{f}^{(\varepsilon)})$  satisfies  $\tilde{\Pi}_z^{(\varepsilon)} \tau = 0$  for every symbol  $\tau$  that contains at least one occurrence of  $\mathcal{E}$ . Therefore, any limiting model  $\tilde{\Pi}$  must satisfy  $\tilde{\Pi}_z \tau = 0$  for such symbols, which is indeed the case for  $\hat{\Pi}$ .

Regarding the symbols  $\tau$  not containing  $\mathcal{E}$ , we see from the definition of  $\mathcal{L}_\varepsilon$  in Section 3.6 that both  $\mathcal{L}_\varepsilon(\xi^{(\varepsilon)})$  and  $\mathcal{L}_0(\xi^{(\varepsilon)})$  act in exactly the same way on these symbols. Furthermore, the map  $\Delta^{\text{Wick}}$  appearing in (5.1) is the same for the constructions of  $M^{(\varepsilon)}$  and  $\tilde{M}^{(\varepsilon)}$ , and the maps  $M_0$  (also appearing in (5.1)) coincide on all elements not containing the symbol  $\mathcal{E}$ . Therefore, we have  $\tilde{\Pi}_z^{(\varepsilon)} \tau = \hat{\Pi}_z^{(\varepsilon)} \tau$  for every  $\tau$  not containing  $\mathcal{E}$ . The claim (including that the models  $\tilde{\mathfrak{M}}_\varepsilon$  converge in  $\mathcal{M}_0$ ) immediately follows from the fact that,  $\tilde{f}^{(\varepsilon)}$  is uniquely determined from  $\tilde{\Pi}^{(\varepsilon)}$  by the condition that our models are admissible and satisfy  $\tilde{f}_z^{(\varepsilon)}(\mathcal{E}_\ell^k(\tau)) = 0$  for every  $\tau$ .  $\square$

*Proof of Proposition 7.1.* By Proposition 7.2,  $h$  is the limit in probability of  $h_\varepsilon$ , where  $h_\varepsilon = \mathcal{R}^{(\varepsilon)} H_\varepsilon$ , with  $H_\varepsilon$  the solution to the fixed point problem associated to the model  $\tilde{\mathfrak{M}}_\varepsilon$  and  $\mathcal{R}^{(\varepsilon)}$  the corresponding reconstruction operator. (Note that  $\tilde{\mathfrak{M}}_\varepsilon$  is a model in  $\mathcal{M}_0$  and the convergence takes place there. As a consequence, we can take an initial condition in  $\mathcal{C}^\eta$  even for  $\varepsilon \neq 0$ .)

However, we know from Proposition 5.9 that  $h_\varepsilon$  is the classical strong solution to the semilinear PDE

$$\partial_t h_\varepsilon = \partial_x^2 h_\varepsilon + \hat{a}_1 (\partial_x h_\varepsilon)^2 + \xi^{(\varepsilon)} - \hat{a}_1 C_0^{(\varepsilon)} - c_\varepsilon, \quad (7.1)$$

where the constant  $c_\varepsilon$  is given by

$$c_\varepsilon = 2\hat{a}_1^3 (4C_2^{(\varepsilon)} + C_3^{(\varepsilon)}).$$

(Recall that the constants  $C_i^{(\varepsilon)}$  were defined in (6.1).) This constant converges to a finite limit of the form  $\hat{a}_1^3 c_0$  with  $c_0 \in \mathbf{R}$  depending in general both on the mollifier  $\varrho$  and the (arbitrary) choice of kernel  $K$  by Theorem 6.5. In particular, a simple application of the chain rule shows that  $Z_\varepsilon = \exp(\hat{a}_1 h_\varepsilon)$  is the mild solution to

$$\partial_t Z_\varepsilon = \partial_x^2 Z_\varepsilon + \hat{a}_1 Z_\varepsilon \xi^{(\varepsilon)} - \hat{a}_1 (\hat{a}_1 C_0^{(\varepsilon)} + c_\varepsilon) Z_\varepsilon. \quad (7.2)$$

It was recently shown in [HP15, Theorem 4.7] (but see also [Hai13]) that, for every  $T > 0$ , the family  $Z_\varepsilon$  converges in probability in  $\mathcal{C}^\eta([0, T] \times S^1)$  to a limit  $Z$  and that, provided that the renormalization constant  $c_\varepsilon$  is suitably chosen (of the form  $\hat{a}_1^3 \hat{c}_0$  for some  $\hat{c}_0$  depending only on the choice of mollifier), this limit is almost surely equal to the solution to the stochastic heat equation (1.3) with  $\lambda = \hat{a}_1$ . This shows that the limit of (7.2) is given by

$$Z = \exp(\hat{a}_0^3 (\hat{c}_0 - c_0)t) Z^{(\hat{a}_1)}.$$

Since we know that  $Z^{(\hat{a}_1)}$  remains strictly positive [BG97], this implies in particular that  $h_\varepsilon - \hat{a}_0^3 (\hat{c}_0 - c_0)t$  converges in probability to  $h_{\text{HC}}^{(\lambda)}$ , thus proving the claim with  $c = \hat{c}_0 - c_0$ . The fact that  $c$  depends only on  $K$  and not on the choice of mollifier  $\varrho$  is a simple consequence of the fact that neither the limiting model  $\mathfrak{M}$  nor the Hopf–Cole solution depend on  $\varrho$ . (But the limiting model  $\mathfrak{M}$  does depend on the choice of  $K$ , this is why there is no ‘canonical’ value for  $c$ .)  $\square$

We are now ready to collect all of these results to prove the main convergence result of this article.

*Proof of Theorem 1.2.* Writing  $\hat{h}_\varepsilon = h_\varepsilon - (\varepsilon^{-1}\hat{\lambda} + c)t$ , we first note that  $\hat{h}_\varepsilon$  solves the equation

$$\partial_t \hat{h}_\varepsilon = \partial_x^2 \hat{h}_\varepsilon + \frac{1}{\varepsilon} F(\sqrt{\varepsilon} \partial_x h_\varepsilon) - \varepsilon^{-1} \hat{\lambda} - c + \xi^{(\varepsilon)}.$$

Define now coefficients  $\hat{a}_k$  implicitly by imposing the identity between polynomials

$$F(x) = \sum_{k=0}^m \hat{a}_k H_{2k}(x, C_0),$$

where  $H_k(x, c)$  denotes the  $k$ th generalized Hermite polynomial as in (5.6). One can check that the coefficients  $\hat{a}_k$  are then given by

$$\hat{a}_k = \frac{1}{k!} \int F^{(k)}(x) \mu_0(dx).$$

As a consequence of Proposition 5.9, it then follows that, provided that the constant  $c$  is suitably chosen and that we set  $\hat{\lambda} = \hat{a}_0$ , one has  $\hat{h}_\varepsilon = \mathcal{R}H$ , where  $H$  solves the fixed point problem (4.20) for the renormalized model  $\mathfrak{M}_\varepsilon$  considered in Theorem 6.1. The (local in time) convergence of  $\hat{h}_\varepsilon$  to a limit  $h$  now follows by combining the convergence of  $\mathfrak{M}_\varepsilon$  given in Theorem 6.1 with Theorem 4.16. The identification of the limit as the Hopf–Cole solution (provided that the constant  $c$  is suitably chosen) is given by Proposition 7.1. Since we know that the Hopf–Cole solutions are global, we immediately obtain convergence over any fixed time interval from the last statement of Theorem 4.16.  $\square$

**7.2. Intermediate disorder regime.** We now prove Theorem 1.1. Let us first consider the special case where  $F$  is a polynomial, so that  $\tilde{F} = 0$ . In this case, we can rewrite the nonlinearity of (1.16) as

$$\sum_{k=0}^{2p-1} a_{p+k} \varepsilon^{k/(2p-1)} \varepsilon^{p+k-1} (\partial_x h_\varepsilon)^{2(p+k)},$$

which suggests that we should set  $F_\varepsilon(x) = \sum_{k=0}^{2p-1} a_{p+k} \varepsilon^{k/(2p-1)} x^{2(p+k)}$  and define coefficients  $\hat{a}_k^{(\varepsilon)}$  as before by

$$\hat{a}_k^{(\varepsilon)} = \frac{1}{k!} \int F_\varepsilon^{(k)}(x) \mu_0(dx).$$

In this case, one has in particular  $\hat{a}_2^{(\varepsilon)} \rightarrow \lambda$ , with  $\lambda$  as in (1.17), as well as  $\hat{a}_0^{(\varepsilon)} = \varepsilon C_\varepsilon$ , with  $C_\varepsilon$  as in (1.17). One also has  $\hat{a}_k^{(\varepsilon)} \rightarrow \hat{a}_k$  for some  $\hat{a}_k$  proportional to  $a_p$  for  $k \leq p$ , and  $\hat{a}_k^{(\varepsilon)} \rightarrow 0$  for  $k \in \{p, \dots, 3p-1\}$ . With these notations at hand,

we consider the fixed point problem

$$\begin{aligned} H_\varepsilon &= \mathcal{P}\mathbf{1}_+ \left( \mathcal{E} + \sum_{j=1}^{3p-1} \hat{a}_j^{(\varepsilon)} \mathcal{Q}_{\leq 0} \hat{\mathcal{E}}^{j-1} (\mathcal{Q}_{\leq 0} (\mathcal{D}H_\varepsilon)^{2j}) \right. \\ &\quad \left. + \varepsilon^{-(3p-1)/(2p-1)} \tilde{F}(\varepsilon^{p/(2p-1)} \mathcal{R}\mathcal{D}H_\varepsilon) \mathbf{1} \right) \\ &\quad + Ph_0^{(\varepsilon)}, \end{aligned} \quad (7.3)$$

where  $\mathcal{R}$  denotes the reconstruction operator. Note now that if  $H_\varepsilon$  solves this fixed point equation and belongs to  $\mathcal{D}_\varepsilon^{\gamma, \eta}$ , then  $\mathcal{D}H_\varepsilon$  is necessarily of the form

$$\mathcal{D}H_\varepsilon = \mathcal{I}'(\mathcal{E}) + U',$$

with  $U' \in \mathcal{D}_\varepsilon^{\gamma, \eta}$  and  $U'$  taking values in the subspace of  $\mathcal{T}$  spanned by  $\mathbf{1}$  and elements with strictly positive homogeneity. In particular, by (4.10) and [Hai14, Definition 6.2],  $\mathcal{R}U'$  is a continuous function such that

$$|(\mathcal{R}U')(t, x)| \lesssim (\varepsilon^2 + |t|)^{(\eta-1)/2} \|H_\varepsilon\|_{\gamma, \eta; \varepsilon}. \quad (7.4)$$

It is also straightforward to show that

$$(\Pi_z^{(\varepsilon)} \mathcal{I}'(\mathcal{E}))(z) = |(K' * \xi^{(\varepsilon)})(z)| \lesssim \varepsilon^{-(1/2)-\kappa}, \quad (7.5)$$

for any  $\kappa > 0$ , uniformly over compact domains. This shows that, for  $\eta > \frac{1}{2} - \kappa$ , the map

$$H_\varepsilon \mapsto \varepsilon^{(1/2)+\kappa} \mathcal{R}\mathcal{D}H_\varepsilon,$$

is locally Lipschitz continuous from  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  into  $\mathcal{C}$  (the space of continuous functions on the compact domain  $D$  endowed with the supremum norm), uniformly over models  $(\Pi, \Gamma) \in \mathcal{M}_\varepsilon$  with  $\|\Pi\|_\varepsilon$  bounded and furthermore satisfying (7.5). Combining this with the fact that  $|\tilde{F}(u) - \tilde{F}(v)| \lesssim |u - v|(|u|^{6p-1} + |v|^{6p-1})$  for  $u$  and  $v$  bounded, we conclude that, provided that  $\kappa < 1/(12p^2)$ , the map

$$H_\varepsilon \mapsto \tilde{F}^{(\varepsilon)}(H_\varepsilon) \stackrel{\text{def}}{=} \varepsilon^{-(3p-1)/(2p-1)} \tilde{F}(\varepsilon^{p/(2p-1)} \mathcal{R}\mathcal{D}H_\varepsilon),$$

is locally Lipschitz continuous from  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  into  $\mathcal{C}$  (the space of continuous functions on the compact domain  $D$  endowed with the supremum norm), with both norm and Lipschitz constant bounded uniformly over  $\varepsilon \in (0, 1]$ ,  $H_\varepsilon$  in bounded balls of  $\mathcal{D}_\varepsilon^{\gamma, \eta}$ , and models in  $\mathcal{M}_\varepsilon$  with bounded norm satisfying (7.5)

for a fixed proportionality constant. As a matter of fact, both the norm and the Lipschitz constant of  $\tilde{F}^{(\varepsilon)}$  are bounded by  $\varepsilon^\theta$  for some  $\theta > 0$ . Since the map  $u \mapsto P * \mathbf{1}_+ u$ , where  $P$  denotes the heat kernel, maps  $\mathcal{C}$  into  $\mathcal{D}_\varepsilon^{\gamma, \eta}$  with norm bounded uniformly in  $\varepsilon$  and behaving like  $T^\theta$  for some  $\theta > 0$ , where  $T$  is the local existence time under consideration, we can proceed as in the proof of Theorem 4.16 to conclude that (7.3) admits local solutions with a local existence time uniform over initial conditions and models as just discussed.

As in the proof of Theorem 4.16, one shows that as  $\varepsilon \rightarrow 0$ , assuming that  $\|\Pi^{(\varepsilon)}; \Pi\|_\varepsilon \rightarrow 0$  for some model  $\Pi$  and that the bound (7.5) holds uniformly over  $\varepsilon \in (0, 1]$ , one has  $\|H_\varepsilon; H\|_{\gamma, \eta; \varepsilon} \rightarrow 0$ , where  $H$  solves the fixed point problem

$$H_\varepsilon = \mathcal{P}\mathbf{1}_+ \left( \varepsilon + \sum_{j=1}^p \hat{a}_j \mathcal{Q}_{\leq 0} \hat{\mathcal{E}}^{j-1} (\mathcal{Q}_{\leq 0} (\mathcal{D}H_\varepsilon)^{2j}) \right) + Ph_0. \quad (7.6)$$

We now conclude exactly as before, noting that if we take for  $\Pi^{(\varepsilon)}$  the model  $\mathfrak{M}_\varepsilon$  considered in Theorem 6.1 then, as a consequence of Proposition 5.9,  $\mathcal{R}H_\varepsilon$  is precisely equal to  $h_\varepsilon - (C_\varepsilon + c_\varepsilon)t$ , for the same constant  $C_\varepsilon$  as in the statement and some constant  $c_\varepsilon$  converging to a limit  $c \in \mathbf{R}$ .

## Acknowledgements

MH gratefully acknowledges financial support from the Leverhulme trust, the Royal Society, the ERC via the consolidator grant 615897:CRITICAL, and the Institute for Advanced Study. JQ gratefully acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada, the I. W. Killam Foundation, and the Institute for Advanced Study.

## Appendix A. A bound on generalized convolutions

In this section we obtain an estimate which allows us to bound the kind of generalized convolutions of kernels appearing in the construction of quite general models built from Gaussian (and other) processes.

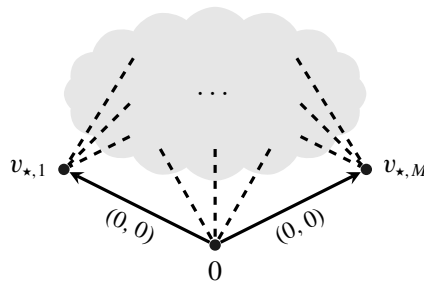
The basic ingredients are the following: a finite directed multigraph  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$  with edges  $e \in \mathbb{E}$  labelled by pairs  $(a_e, r_e) \in \mathbf{R}_+ \times \mathbf{Z}$ , and kernels  $K_e: \mathbf{R}^d \setminus \{0\} \rightarrow \mathbf{R}$  which are compactly supported in the ball of radius 1 around the origin. By multigraph we mean that we allow (a finite number of) multiple edges between vertices. However, we will not allow edges from a vertex to itself (loops). We will always assume that every vertex has either an outgoing or incoming edge. The exponent  $a_e$  describes the singularity of the kernel  $K_e$  at

the origin in the sense that we assume that, for every  $p > 0$  and every edge  $e \in \mathbb{E}$ , the quantity  $\|K_e\|_{a_e, p}$  is finite, where

$$\|K\|_{\alpha, p} \stackrel{\text{def}}{=} \sup_{\substack{\|x\|_S \leq 1 \\ |k|_S \leq p}} \|x\|_S^{\alpha+|k|_S} |D^k K(x)| < \infty. \quad (\text{A.1})$$

The constant  $r_e$  will be used to allow for a renormalization of the singularity. The kernels are otherwise assumed to be smooth. If  $r_e < 0$ , then we will in addition be given a collection of real numbers  $\{I_{e,k}\}_{|k|_S < |r_e|}$  used to identify a Schwartz distribution associated to the singularity (see (A.5)).

We will always consider the situation where  $\mathbb{G}$  contains a finite number  $M \geq 1$  (typically  $M = 2$ ) of distinguished edges  $e_{\star,1}, \dots, e_{\star,M}$  connecting a distinguished vertex  $0 \in \mathbb{V}$  to  $M$  distinct vertices  $v_{\star,1}, \dots, v_{\star,M}$ , and all with label  $(a_e, r_e) = (0, 0)$ . In other words, the graphs we consider will always be of the following type:



(In particular, some of the nondistinguished vertices may connect to 0.) Here, the integer  $M$  will denote the moment we are currently controlling and the distinguished edges will represent test functions. We will use the notation  $\mathbb{V}_\star \subset \mathbb{V}$  for the set consisting of the special vertex 0, plus the vertices  $v_{\star,i}$ , and we write  $\mathbb{V}_0 = \mathbb{V} \setminus \{0\}$ .

Given a directed edge  $e \in \mathbb{E}$ , we write  $e_\pm$  for the two vertices so that  $e = (e_-, e_+)$  is directed from  $e_-$  to  $e_+$ . In cases where there is more than one edge connecting  $e_-$  to  $e_+$  we will always assume that at most one can have nonzero renormalization  $r_e$ , and in that case  $r_e$  must be positive. Then we may identify the multigraph with a graph  $(\mathbb{V}, \hat{\mathbb{E}})$  where the multiedges from  $e_-$  to  $e_+$  are merged into one single edge whose label  $(\hat{a}_e, r_e)$  is simply the sum of the labels on the original multiedges. The rest of the assumptions are most easily stated in terms of these new labels on the resulting directed graph  $(\mathbb{V}, \hat{\mathbb{E}})$ , although the application will be to the generalized convolution on the original graph. We will also sometimes make the abuse of notation that identifies  $e$  with the set

$\{e_-, e_+\}$  even though our edges are directed. A subset  $\tilde{\mathbb{V}} \subset \mathbb{V}$  has *outgoing edges*  $\mathbb{E}^\uparrow(\tilde{\mathbb{V}}) = \{e \in \mathbb{E} : e \cap \tilde{\mathbb{V}} = e_-\}$ , *incoming edges*  $\mathbb{E}^\downarrow(\tilde{\mathbb{V}}) = \{e \in \mathbb{E} : e \cap \tilde{\mathbb{V}} = e_+\}$ , *internal edges*  $\mathbb{E}_0(\tilde{\mathbb{V}}) = \{e \in \mathbb{E} : e \cap \tilde{\mathbb{V}} = e\}$ , and *incident edges*  $\mathbb{E}(\tilde{\mathbb{V}}) = \{e \in \mathbb{E} : e \cap \tilde{\mathbb{V}} \neq \emptyset\}$ . We will also use  $\mathbb{E}_+(\tilde{\mathbb{V}}) = \{e \in \mathbb{E}(\tilde{\mathbb{V}}) : r_e > 0\}$  to denote the edges with positive renormalization,  $\mathbb{E}_+^\uparrow = \mathbb{E}_+ \cap \mathbb{E}^\uparrow$  and  $\mathbb{E}_+^\downarrow = \mathbb{E}_+ \cap \mathbb{E}^\downarrow$ .

ASSUMPTION A.1. The resulting directed graph  $(\mathbb{V}, \hat{\mathbb{E}})$  with labels  $(\hat{a}_e, r_e)$  satisfies: no edge with  $r_e \neq 0$  connects two elements in  $\mathbb{V}_*$  and  $0 \in e \Rightarrow r_e = 0$ ; no more than one edge with negative renormalization  $r_e < 0$  may be incident to the same vertex; and

- (1) For all  $e \in \hat{\mathbb{E}}$ , one has  $\hat{a}_e + (r_e \wedge 0) < |\mathfrak{s}|$  where  $|\mathfrak{s}| = \sum_i \mathfrak{s}_i$ ;
- (2) for every subset  $\tilde{\mathbb{V}} \subset \mathbb{V}_0$  of cardinality at least 3,

$$\sum_{e \in \hat{\mathbb{E}}_0(\tilde{\mathbb{V}})} \hat{a}_e < (|\tilde{\mathbb{V}}| - 1)|\mathfrak{s}|; \quad (\text{A.2})$$

- (3) for every subset  $\tilde{\mathbb{V}} \subset \mathbb{V}$  containing 0 of cardinality at least 2,

$$\sum_{e \in \hat{\mathbb{E}}_0(\tilde{\mathbb{V}})} \hat{a}_e + \sum_{e \in \hat{\mathbb{E}}_+^\uparrow(\tilde{\mathbb{V}})} (\hat{a}_e + r_e - 1) - \sum_{e \in \hat{\mathbb{E}}_+^\downarrow(\tilde{\mathbb{V}})} r_e < (|\tilde{\mathbb{V}}| - 1)|\mathfrak{s}|; \quad (\text{A.3})$$

- (4) for every nonempty subset  $\tilde{\mathbb{V}} \subset \mathbb{V} \setminus \mathbb{V}_*$ ,

$$\sum_{e \in \hat{\mathbb{E}}(\tilde{\mathbb{V}}) \setminus \hat{\mathbb{E}}_+^\downarrow(\tilde{\mathbb{V}})} \hat{a}_e + \sum_{e \in \hat{\mathbb{E}}_+^\uparrow(\tilde{\mathbb{V}})} r_e - \sum_{e \in \hat{\mathbb{E}}_+^\downarrow(\tilde{\mathbb{V}})} (r_e - 1) > |\tilde{\mathbb{V}}||\mathfrak{s}|. \quad (\text{A.4})$$

Next we describe the renormalization procedure. If  $r_e < 0$ , then, in a way reminiscent of [BP57, Hep69, Zim69], we associate to  $K_e$  the distribution,

$$(\mathcal{R}K_e)(\varphi) = \int K_e(x) \left( \varphi(x) - \sum_{|k|_{\mathfrak{s}} < |r_e|} \frac{x^k}{k!} D^k \varphi(0) \right) dx + \sum_{|k|_{\mathfrak{s}} < |r_e|} \frac{I_{e,k}}{k!} D^k \varphi(0). \quad (\text{A.5})$$

Note that Assumption A.1(1) and (A.1) imply that the integral in the definition of  $\mathcal{R}K_e$  converges, so that this definition actually makes sense. In some situations, it is natural to choose the constants  $I_{e,k}$  to vanish, but this is not always the case. For example, this allows us to deal with ‘kernels’ for which  $K = 0$ , representing Dirac’s delta-function. It also allows to deal with the fact that the choice of  $K_e$

sometimes contains some arbitrariness (in our case this might be an arbitrary cutoff of the heat kernel), so that setting  $I_{e,k} = 0$  is itself the consequence of an arbitrary choice.

Of course, if  $\int |K_e(x)| |x|^k dx < \infty$  for  $|k|_s < |r_e|$  and  $I_{e,k} = \int K_e(x) x^k dx$ , then one just has  $(\mathcal{R}K_e)(\varphi) = \int K_e(x) \varphi(x) dx$ . For  $r_e \geq 0$ , we just define  $(\mathcal{R}K_e)(\varphi) = \int K_e(x) \varphi(x) dx$ .

Equation (A.5) defines a distributional ‘kernel’  $\hat{K}_e$  for  $r_e < 0$  acting on smooth  $\varphi$  on  $\mathbf{R}^d \times \mathbf{R}^d$  by

$$\hat{K}_e(\varphi) \stackrel{\text{def}}{=} \frac{1}{2} \int \mathcal{R}K_e(\varphi_z) dz,$$

where  $\varphi_z(\bar{z}) \stackrel{\text{def}}{=} \varphi((z + \bar{z})/2, (z - \bar{z})/2)$ . This may appear to be a slightly strange definition at first sight, but if  $\mathcal{R}K_e$  happens to be represented by a measurable function  $\overline{\mathcal{R}K_e}$ , then this definition is equivalent to

$$\hat{K}_e(\varphi) = \iint \overline{\mathcal{R}K_e}(x_{e_+} - x_{e_-}) \varphi(x_{e_+}, x_{e_-}) dx_{e_-} dx_{e_+}.$$

For  $r_e \geq 0$ , we similarly define

$$\hat{K}_e(x_{e_-}, x_{e_+}) = K_e(x_{e_+} - x_{e_-}) - \sum_{|j|_s < r_e} \frac{x_{e_+}^j}{j!} D^j K_e(-x_{e_-}). \quad (\text{A.6})$$

Then it is natural to say that a sequence of such collections of kernels  $K_m \rightarrow K$  if they satisfy (A.1) uniformly in  $m$ , and, for each  $e$  with  $r_e \geq 0$ ,  $D^j K_{e,m} \rightarrow D^j K_e$  for  $|j|_s < r_e$  pointwise on  $x \in \mathbf{R}^d \setminus \{0\}$  as  $m \rightarrow \infty$ , and for each  $e$  with  $r_e < 0$ ,  $K_{e,m} \rightarrow K_e$  pointwise and  $I_{e,k,m} \rightarrow I_{e,k}$ ,  $|k|_s < |r_e|$ .

REMARK A.2. In principle, one may encounter situations where more sophisticated renormalization procedures are required. For the purpose of the present article however, the procedure described here is sufficient.

For a smooth test function  $\varphi$ , let  $\varphi_\lambda(x) = \lambda^{-|s|} \varphi(x/\lambda)$ . The key quantity of interest is the generalized convolution

$$\mathcal{I}^{\mathbb{G}}(\varphi_\lambda, K) \stackrel{\text{def}}{=} \int_{(\mathbf{R}^d)^{\mathbb{V}_0}} \prod_{e \in \mathbb{E}} \hat{K}_e(x_{e_-}, x_{e_+}) \prod_{i=1}^M \varphi_\lambda(x_{v_{*,i}}) dx. \quad (\text{A.7})$$

It is not obvious in general whether the right hand side of (A.7) even makes sense, but actually it is not so hard to see that our conditions imply that in the



situation where the kernels  $K$  are given by smooth functions, the distributions  $\mathcal{R}K_e$ ,  $r_e < 0$  also act only on smooth functions. The fact that it does still make sense in general is part of the following statement, which is the main result of this section.

**THEOREM A.3.** Let  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$  be a finite directed multigraph with labels  $\{a_e, r_e\}_{e \in \mathbb{E}}$  and kernels  $\{K_e\}_{e \in \mathbb{E}}$  with the resulting graph satisfying Assumption A.1 and its preamble. Then, there exist  $C, p < \infty$  depending only on the structure of the graph  $(\mathbb{V}, \mathbb{E})$  and the labels  $r_e$  such that

$$\mathcal{I}^{\mathbb{G}}(\varphi_\lambda, K) \leq C \lambda^{\tilde{\alpha}} \prod_{e \in \mathbb{E}} \|K_e\|_{a_e, p}, \quad (\text{A.8})$$

for  $0 < \lambda \leq 1$ , where

$$\tilde{\alpha} = |\mathbb{S}||\mathbb{V} \setminus \mathbb{V}_*| - \sum_{e \in \mathbb{E}} a_e.$$

In particular, the generalized convolution in (A.7) is well defined in the sense that if  $K_m \rightarrow K$ , then  $\mathcal{I}^{\mathbb{G}}(\varphi_\lambda, K_m)$  converges to a limit  $\mathcal{I}^{\mathbb{G}}(\varphi_\lambda, K)$  that is independent of the approximating sequences  $K_m$ .

Note in particular, that the bound (A.8) for genuine distributions, that is, kernels  $K_e$  with nonintegrable singularities at 0 and  $r_e < 0$ , follows immediately once we prove the bound for regularizations of the kernels, but with the norms on the right hand side independent of the regularization. This has the consequence that within the proof, we can assume without loss of generality that all the kernels are smooth on all of  $\mathbf{R}^d$ . The theorem will be proved in subsections A.1–A.6.

**A.1. Decomposition.** To simplify notations in what follows, we will start by enhancing the set of edges in our graph to include any  $(v, w) \in \mathbb{V}^2$  for which there is not already one, or several, edges in  $\mathbb{E}$ . To all such new directed edges we simply assign the kernel  $\hat{K}_{(v,w)} \equiv 1$ , so that, since every vertex of the original graph had either an incoming or outgoing edge, (A.7) is unaffected, and the fact that these new kernels do not have compact support is irrelevant. These new edges necessarily come with  $a_e = r_e = 0$ . We will abuse notation somewhat by henceforth referring to this enhanced graph as  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ .

Now define a sequence of kernels  $\{K_e^{(n)}\}_{n \geq 0}$  through the following.

**LEMMA A.4.** If  $K_e$  are as above, then there exist  $\{K_e^{(n)}\}_{n \geq 0}$  satisfying:

- (1)  $K_e(x) = \sum_{n \geq 0} K_e^{(n)}(x)$  for all  $x \neq 0$ ;
- (2)  $(\mathcal{R}K_e)(\varphi) = \sum_{n \geq 0} \int K_e^{(n)}(x) \varphi(x) dx$  for smooth test functions  $\varphi$ ;

(3)  $K_e^{(n)}$  is supported in the annulus  $2^{-(n+2)} \leq \|x\|_s \leq 2^{-n}$ ;

(4) for some  $C < \infty$

$$\sup_{|k| \leq p, n \geq 0} 2^{-(a_e + |k|_s)n} |D^k K_e^{(n)}(x)| \leq C \|K_e\|_{a_e, p}; \quad (\text{A.9})$$

(5) if  $r_e < 0$ , then  $\int P(x) K_e^{(n)}(x) dx = 0$  for all  $n > 0$  and all polynomials  $P$  with scaled degree strictly less than  $|r_e|$ .

Note that we do not impose 5. for  $n = 0$  which is why it can be consistent with 2.

*Proof.* We first treat the case  $r_e \geq 0$ . Let  $\psi: \mathbf{R} \rightarrow [0, 1]$  be a smooth function supported on  $[3/8, 1]$  and such that  $\sum_{n \in \mathbf{Z}} \psi(2^n x) = 1$  for every  $x \neq 0$ , and let

$$\Psi^{(n)}(x) = \psi(2^n \|x\|_s), \quad (\text{A.10})$$

so that  $\Psi^{(n)}$  is supported in  $2^{-(n+2)} \leq \|x\|_s \leq 2^{-n}$ , satisfies (A.9) with  $a_e$  replaced by 0, and sums up to 1. We also use the shorthand  $\Psi^{(-)}(x) = \sum_{n \leq 0} \Psi^{(n)}(x)$ .

Let  $K_e^{(0)}(x) = \Psi^{(-)}(x) K_e(x)$  and  $K_e^{(n)}(x) = \Psi^{(n)}(x) K_e(x)$  for  $n > 0$ . As a consequence of (A.1), and the fact that  $|D^k \Psi^{(n)}(x)| \lesssim \|x\|_s^{-|k|_s}$ , it is then straightforward to verify that  $K_e^{(n)}$  does indeed satisfy the claimed properties.

In the case  $r_e < 0$ , the situation is a little less straightforward since then 2. does not follow from 1. and we also want to impose 5. In order to achieve this, we first note that it is possible to find smooth functions  $\eta_k: \mathbf{R}^d \rightarrow \mathbf{R}$  which are supported in the annulus  $\{x : \|x\|_s \in [1/4, 1/2]\}$  and are such that  $\int x^\ell \eta_k(x) dx = \delta_{k,\ell}$  for every  $\ell$  with  $|\ell|_s < |r_e|$ . We also set

$$\eta_k^{(n)}(x_1, \dots, x_d) = 2^{n(|s| + |k|_s)} \eta_k(2^{n s_1} x_1, \dots, 2^{n s_d} x_d).$$

We then set  $I_{e,k}^{(0)} \stackrel{\text{def}}{=} I_{e,k} - \int x^k \Psi^{(-)}(x) K_e(x) dx$ ,

$$K_e^{(0)}(x) \stackrel{\text{def}}{=} \Psi^{(-)}(x) K_e(x) + \sum_{|k|_s < |r_e|} \eta_k^{(0)}(x) I_{e,k}^{(0)},$$

and recursively for  $n > 0$ ,  $I_{e,k}^{(n)} \stackrel{\text{def}}{=} I_{e,k}^{(n-1)} - \int x^k \Psi^{(n)}(x) K_e(x) dx$ ,

$$K_e^{(n)}(x) \stackrel{\text{def}}{=} \Psi^{(n)}(x) K_e(x) + \sum_{|k|_s < |r_e|} (\eta_k^{(n)}(x) I_{e,k}^{(n)} - \eta_k^{(n-1)}(x) I_{e,k}^{(n-1)}).$$

With this definition, it is then straightforward to verify that 1. is satisfied due to the fact that the additional terms form a telescopic sum (with only finitely

many nonzero terms), 4. is satisfied since  $\Psi^{(n)}$  satisfies (A.9) with  $a_e$  replaced by 0, and, as a consequence of the definition of the coefficients  $I_{e,k}^{(n)}$  one has  $\int K_e^{(n)}(x)x^k dx = I_{e,k}\delta_{n,0}$  for  $|k|_s < |r_e|$ . which proves 5. Finally, to prove 2. we have, for each  $n > 0$ ,  $\sum_{j=0}^n \int K_e^{(n)}(x)\varphi(x) dx$  is given by

$$\begin{aligned} & \int \sum_{j \leq n} \Psi^{(j)}(x) K_e(x) \left( \varphi(x) - \sum_{|k|_s < |r_e|} x^k \int \eta_k^{(n)}(y) \varphi(y) dy \right) dx \\ & + \sum_{|k|_s < |r_e|} I_{e,k} \int \eta_k^{(n)}(y) \varphi(y) dy, \end{aligned}$$

which for smooth test functions converges to (A.5) as  $n \rightarrow \infty$ .  $\square$

DEFINITION A.5. For  $\mathbf{n} \in \mathbf{N}^3$  define  $\hat{K}_e^{(\mathbf{n})}(x, y)$  as follows: if  $r_e \leq 0$ , then  $\hat{K}_e^{(\mathbf{n})} = 0$  unless  $\mathbf{n} = (k, 0, 0)$  in which case  $\hat{K}_e^{(\mathbf{n})}(x, y) = K_e^{(k)}(y - x)$  with  $K_e^{(k)}$  given by Lemma A.4; if  $r_e > 0$ , then

$$\begin{aligned} \hat{K}_e^{(k,p,m)}(x, y) &= \Psi^{(k)}(y - x) \Psi^{(p)}(x) \Psi^{(m)}(y) \\ &\times \left( K_e(y - x) - \sum_{|j|_s < r_e} \frac{y^j}{j!} D^j K_e(-x) \right), \end{aligned} \quad (\text{A.11})$$

where the functions  $\Psi^{(k)}$  are defined in (A.10).

For  $\mathbf{n}: \mathbb{E} \rightarrow \mathbf{N}^3$ , let

$$\hat{K}^{(\mathbf{n})}(x) = \prod_{e \in \mathbb{E}} \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}) \quad (\text{A.12})$$

so that if  $K_e$  are smooth on all of  $\mathbf{R}^d$ ,

$$\mathcal{I}^{\mathbb{G}}(\varphi_\lambda, K) = \sum_{\mathbf{n}} \int_{(\mathbf{R}^d)^{\mathbb{V}_0}} \hat{K}^{(\mathbf{n})}(x) \prod_{i=1}^M \varphi_\lambda(x_{v_{*,i}}) dx.$$

For  $\lambda \in (0, 1]$ , let

$$\mathcal{N}_\lambda \stackrel{\text{def}}{=} \{\mathbf{n}: \mathbb{E} \rightarrow \mathbf{N}^3 : 2^{-|\mathbf{n}_{e_{*,i}}|} \leq \lambda, i = 1, \dots, M\}$$

where  $e_{*,i} = (0, v_{*,i})$ . Let

$$\mathcal{I}_\lambda^{\mathbb{G}}(K) \stackrel{\text{def}}{=} \sum_{\mathbf{n} \in \mathcal{N}_\lambda} \int_{(\mathbf{R}^d)^{\mathbb{V}_0}} \hat{K}^{(\mathbf{n})}(x) dx. \quad (\text{A.13})$$

REMARK A.6. The main reason to add all the extra edges with  $K_e = 1$  is that  $\mathbf{n}$  now completely determines the distance (up to a factor 4) between any two coordinates  $x_v$  and  $x_w$  of  $x \in (\mathbf{R}^d)^{\mathbb{V}_0}$ .

Theorem A.3 follows from

LEMMA A.7. *Under the same assumptions as Theorem A.3, there exist  $C$ ,  $p < \infty$  depending only on the structure of the graph  $(\mathbb{V}, \mathbb{E})$  and the labels  $r_e$  such that*

$$|\mathcal{I}_\lambda^{\mathbb{G}}(K)| \leq C \lambda^\alpha \prod_{e \in \mathbb{E}} \|K_e\|_{a_e; p}, \quad \lambda \in (0, 1], \quad (\text{A.14})$$

where  $\alpha = |\mathbb{S}||\mathbb{V}_0| - \sum_{e \in \mathbb{E}} a_e$ .

To see that Lemma A.7 implies Theorem A.3 for smooth kernels, we use the fact that the rescaled test function can be viewed as just another kernel  $K_{e_{*,i}}(v_{*,i}) := \varphi_\lambda(v_{*,i})$  with  $a_e = 0$  and  $\|K_{e_{*,i}}\|_{a_e; p} = \lambda^{-|\mathbb{S}|}$ .

To see that it suffices to prove Theorem A.3 for smooth kernels, we argue as follows. Given a labelled graph  $\mathbb{G}$ , let  $p$  be given by the theorem. Given singular kernels  $K_e$  with  $\|K_e\|_{a_e; p} < \infty$ ,  $e \in \mathbb{E}$ , let  $K_{e,m}$  be smooth kernels with  $\|K_{e,m} - K_e\|_{a_e; p} \rightarrow 0$  as  $m \rightarrow \infty$  for each  $e$ . By the multilinearity it is not hard to see that the real numbers  $\mathcal{I}_\lambda^{\mathbb{G}}(\varphi_\lambda, K_m)$ ,  $m = 1, 2, \dots$  form a Cauchy sequence, and therefore have a unique limit, which, in addition, satisfies the bound (A.8).

The lemma will be proved in Subsections A.2–A.6. Throughout this section, the symbol  $\sim$  denotes a bound from above and below, with proportionality constants that only depend on  $|\mathbb{V}|$ . Note that all constructions are finite so, for example, the constants appearing in inductive proofs are allowed to get worse at each stage, and no effort has been made to optimize the dependence on the size of the graph  $\mathbb{G}$ . Note that we can reduce ourselves to the case where all  $\|K_e\|_{a_e; p} = 1$  by multilinearity, so we will not follow these norms in the sequel.

**A.2. Multiscale clustering.** It turns out to be convenient to think of the integral in (A.13) as over  $x \in (\mathbf{R}^d)^{\mathbb{V}}$ , with  $x_0 = 0$  and we will use this convention throughout the proof. Since our kernels are smooth, the set of  $x \in (\mathbf{R}^d)^{\mathbb{V}}$  where any two different  $\|x_v - x_w\|_{\mathbb{S}}$  coincide can be ignored in the integral in (A.13). To other points  $x \in (\mathbf{R}^d)^{\mathbb{V}}$  we will associate a labelled rooted binary tree  $T$  whose leaves are the  $v \in \mathbb{V}$ .

We will use the terminology *node* instead of *vertex* to distinguish the nodes of this tree from the vertices  $\mathbb{V}$  of the original graph, and denote them by  $v$ ,  $\omega$ , and so on. A *leaf* is a node of degree 1. An *inner node* is one of degree at least 2.

A rooted tree comes with a partial order,  $v \geq \omega$  means that  $\omega$  belongs to the shortest path connecting  $v$  to the root. In genealogical terms,  $\omega$  is an ancestor of  $v$ . For any two nodes  $v$  and  $\omega$ , we write  $v \wedge \omega$  for the unique node such that for any node  $v$  satisfying  $v \leq v$  and  $v \leq \omega$ , one necessarily has  $v \leq (v \wedge \omega)$ , that is,  $v \wedge \omega$  is the most recent common ancestor of  $v$  and  $\omega$ . We will furthermore impose that every inner node has exactly two descendants, that only the inner nodes are labelled, by natural numbers, and that the labelling  $\ell$  of the inner nodes respects the partial order in the sense that  $\ell_v \geq \ell_\omega$  whenever  $v \geq \omega$ . Note that the leaves of the tree will sometimes be denoted  $v, w$  since they are also elements of  $\mathbb{V}$ .

The way the tree is constructed is as follows: first consider the complete undirected weighted graph with vertices  $v \in \mathbb{V}$ , and edge weight  $\|x_v - x_w\|_s$  assigned to the edge  $(v, w)$ ,  $v, w \in \mathbb{V}$ . A minimal spanning tree can be constructed, for example, by Kruskal's algorithm [Kru56]: choose first the edge of minimal weight, then successively add the edge with the smallest weight which is not in the tree already, as long as adding it does not create a loop, in which case, it is skipped and we attempt to add the next smallest weight. Since the edge weights can be strictly ordered, there is no ambiguity in this definition. The binary tree  $T$  with leaves  $v \in \mathbb{V}$  simply records the order in which edges were added to the minimal spanning tree: at the stage when the edge  $(v, w)$  is added to the minimal spanning tree, the branch containing  $v$  is joined to the branch containing  $w$ .

Now for each node  $v$  we let

$$\ell_v = \max_{v \wedge w = v} \lfloor -\log_2 \|x_v - x_w\|_s \rfloor.$$

From the construction, if  $v \geq \omega$ , then  $\ell_v \geq \ell_\omega$ .

Given a set of vertices  $\mathbb{V}$ , denote by  $\mathbb{T}(\mathbb{V})$  the set of rooted labelled binary trees  $(T, \ell)$  as above, with an order preserving labelling  $\ell$ , which have  $\mathbb{V}$  as their set of leaves. From the construction, a generic  $x \in (\mathbf{R}^d)^\mathbb{V}$  corresponds to an element  $(T, \ell)$  of  $\mathbb{T}(\mathbb{V})$ . The downside of course is that we can only partially read off the edge lengths  $\|x_v - x_w\|_s$  from  $(T, \ell)$ . More precisely, for any two leaves  $v, w \in \mathbb{V}$ , one has  $\|x_v - x_w\|_s \sim 2^{-\ell_{v \wedge w}}$ , however the constants of proportionality can be quite poor. In particular, it is not hard to see that

$$2^{-\ell_{v \wedge w}} \leq \|x_v - x_w\|_s \leq |\mathbb{V}| 2^{-\ell_{v \wedge w}}, \quad (\text{A.15})$$

and that the upper bound cannot really be improved (for example, place the points colinearly, with the largest gap at one end). In applications such as cladograms, this renders such constructions essentially worthless, however, in our application, it only means that the resulting constant  $C$  on the right hand side of (A.14) will depend badly on the size of the vertex set  $\mathbb{V}$ . Since in

any subcritical stochastic PDE there are only finitely many universal objects to control, the resulting bound suffices for our purposes.

**DEFINITION A.8.** For  $c = \log |\mathbb{V}| + 2$ , let  $\mathcal{N}(T, \ell)$  consist of all functions  $\mathbf{n}: \mathbb{E} \rightarrow \mathbf{N}^3$  such that for every edge  $e = (v, w)$  with  $r_e \leq 0$ , one has  $\mathbf{n}_e = (k, 0, 0)$  with  $|k - \ell_{v \wedge w}| \leq c$ , and for every edge  $e = (v, w)$  with  $r_e > 0$ , one has  $\mathbf{n}_e = (k, p, m)$  with  $|k - \ell_{v \wedge w}| \leq c$ ,  $|p - \ell_{v \wedge 0}| \leq c$ , and  $|m - \ell_{w \wedge 0}| \leq c$ .

If  $\mathbf{n}: \mathbb{E} \rightarrow \mathbf{N}^3$  is such that  $\int_{(\mathbf{R}^d)^\mathbb{V}} \hat{K}^{(\mathbf{n})}(x) dx$  is nonvanishing, then the support of  $\hat{K}^{(\mathbf{n})}$  is nonempty. From (A.13),  $x \in \mathbf{R}^\mathbb{V}$  is in that support only if it belongs to the support of  $\hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+})$  for every  $e \in \mathbb{E}$ . Let  $(T, \ell) \in \mathbb{T}(\mathbb{V}_0)$  be the tree associated to  $x \in \mathbf{R}^\mathbb{V}$ . If  $r_e \leq 0$ , then from Definition A.5, we have that  $\mathbf{n}_e = (k, 0, 0)$  and  $\hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}) = K^{(k)}(x_{e_+} - x_{e_-}) \neq 0$ . From 3 of Lemma A.4 we have  $\|x_{e_+} - x_{e_-}\|_s \in [2^{-k-2}, 2^{-k}]$ , and then from (A.15), we have  $|k - \ell_{e_- \wedge e_+}| \leq c$ . If  $r_e > 0$ , the kernel  $\hat{K}_e^{(\mathbf{n}_e)}$  with  $\mathbf{n}_e = (k, p, m)$  is given by (A.11), so that for  $x$  to belong to its support we must have  $\|x_{e_+} - x_{e_-}\|_s \in [2^{-k-2}, 2^{-k}]$ ,  $\|x_{e_+}\|_s \in [2^{-m-2}, 2^{-m}]$ , as well as  $\|x_{e_-}\|_s \in [2^{-p-2}, 2^{-p}]$ , which in the same way implies  $|k - \ell_{e_- \wedge e_+}| \leq c$ ,  $|p - \ell_{e_- \wedge 0}| \leq c$ , and  $|m - \ell_{e_+ \wedge 0}| \leq c$ . Hence we have

**LEMMA A.9.** For every  $\mathbf{n}: \mathbb{E} \rightarrow \mathbf{N}^3$  such that  $\int_{(\mathbf{R}^d)^\mathbb{V}_0} \hat{K}^{(\mathbf{n})}(x) dx$  from (A.13) is nonvanishing, there exists an element  $(T, \ell) \in \mathbb{T}(\mathbb{V})$  with  $\mathbf{n} \in \mathcal{N}(T, \ell)$ .  $\square$

Denote by  $\mathbb{T}_\lambda(\mathbb{V})$ , the subset of those labelled trees in  $\mathbb{T}(\mathbb{V})$  with the property that  $2^{-\ell_{v \wedge w}} \leq \lambda$  for any two leaves  $v, w \in \mathbb{V}_\star$  (as defined on page 87). As a consequence of Lemma A.9, we can turn the sum over  $\mathcal{N}_\lambda$  appearing in the definition of  $\mathcal{I}_\lambda^\mathbb{G}(K)$  into a sum over  $\mathbb{T}_\lambda(\mathbb{V})$ :

$$|\mathcal{I}_\lambda^\mathbb{G}(K)| \lesssim \sum_{(T, \ell) \in \mathbb{T}_\lambda(\mathbb{V})} \sum_{\mathbf{n} \in \mathcal{N}(T, \ell)} \left| \int_{(\mathbf{R}^d)^\mathbb{V}} \hat{K}^{(\mathbf{n})}(x) dx \right|. \quad (\text{A.16})$$

In order to bound the right hand side we will use the following construction. Consider a rooted binary tree  $T$  with a fixed distinguished inner node  $v_\star$  (in particular it has at least one inner node). We will denote by  $T^\circ$  the set of inner nodes of  $T$ . Since the tree is binary, every node of the subtree  $T^\circ \subset T$  has exactly two children (in  $T$ ), so that  $T^\circ$ , together with its partial order, actually determines the full tree  $T$ . We then consider the set  $\mathcal{N}_\lambda(T^\circ)$  of all integer labelings  $\ell: T^\circ \rightarrow \mathbf{N}$  which preserve the partial order of the tree  $T^\circ$  as above and are such that  $2^{-\ell_{v_\star}} \leq \lambda$ . Finally, given a function  $\eta: T^\circ \rightarrow \mathbf{R}$ , we write

$$\mathcal{I}_\lambda(\eta) = \sum_{\ell \in \mathcal{N}_\lambda(T^\circ)} \prod_{v \in T^\circ} 2^{-\ell_v \eta_v}.$$

Setting  $|\eta| = \sum_{v \in T^\circ} \eta_v$ , we then have the following bound

LEMMA A.10. Assume that  $\eta$  satisfies the following two properties:

- (1) For every  $v \in T^\circ$ , one has  $\sum_{v \geq v} \eta_v > 0$ .
- (2) For every  $v \in T^\circ$  such that  $v \leq v_*$ , one has  $\sum_{v \not\geq v} \eta_v < 0$ , provided that this sum contains at least one term.

Then, one has  $\mathcal{I}_\lambda(\eta) \lesssim \lambda^{|\eta|}$ , uniformly over  $\lambda \in (0, 1]$ .

REMARK A.11. Since the order on  $T^\circ$  is only partial,  $v \not\geq v$  is different from  $v < v$ . The latter would only consider the nodes between  $v$  and the root, while the former also includes the subtrees dangling from these nodes. Note also that the second condition above is empty (and therefore automatically satisfied) in the special case when  $v_*$  also happens to be the root.

REMARK A.12. As will be evident from the proof, the first condition is necessary for  $\mathcal{I}_\lambda(\eta)$  to even be finite. Regarding the second condition, if it fails, then for every  $v$  with  $v_* \geq v$  such that  $\sum_{v \not\geq v} \eta_v = \alpha > 0$ , the upper bound for  $\mathcal{I}_\lambda(\eta)$  is larger by a factor  $\lambda^{-\alpha}$ . If  $\sum_{v \not\geq v} \eta_v = 0$ , one loses a factor  $|\log \lambda|$ .

*Proof of Lemma A.10.* The proof goes by induction on the size of  $T^\circ$ . If  $|T^\circ| = 1$ , it consists of only the node  $v_*$ . Condition 1 implies that  $\eta_{v_*} > 0$  and one has  $\mathcal{I}_\lambda(\eta) = \sum_{2^{-\ell} \leq \lambda} 2^{-\ell \eta_{v_*}} \sim \lambda^{\eta_{v_*}} = \lambda^{|\eta|}$  as required.

If  $|T^\circ| > 1$ , we distinguish between two different cases. In the first case,  $T^\circ$  contains at least one extremal node  $v$  which is different from the distinguished node  $v_*$ . (We call leaves of  $T^\circ$  ‘extremal nodes’ in order not to confuse them with the leaves of  $T$ .) In this case, one has  $\eta_v > 0$  by the first assumption (since  $v$  is extremal, the only  $v$  with  $v \geq v$  is  $v$  itself). Denote now by  $\tilde{T}^\circ$  the tree obtained by erasing the leaf  $v$  and by  $\tilde{\eta}: \tilde{T}^\circ \rightarrow \mathbf{R}$  the function obtained by setting  $\tilde{\eta}_v = \eta_v$  for every node  $v$  which is not the parent  $v_\uparrow$  of  $v$ . We also set  $\tilde{\eta}_{v_\uparrow} = \eta_{v_\uparrow} + \eta_v$ , which ensures that  $\tilde{\eta}$  still satisfies conditions 1 and 2. One then has

$$\begin{aligned} \mathcal{I}_\lambda(\eta) &= \sum_{\ell \in \mathcal{N}_\lambda(T^\circ)} \prod_{\omega \in T^\circ} 2^{-\ell_\omega \eta_\omega} = \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} \sum_{m \geq \ell_{v_\uparrow}} 2^{-m \eta_v} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \eta_\omega} \\ &\sim \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} 2^{-\ell_{v_\uparrow} \eta_v} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \eta_\omega} = \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \tilde{\eta}_\omega} = \mathcal{I}_\lambda(\tilde{\eta}). \end{aligned}$$

By the induction hypothesis the required upper and lower bounds hold.

On the other hand, it may happen that the only extremal node of  $T^\circ$  is  $v_*$  itself. In this case, the tree  $T^\circ$  has a total order and, if  $|T^\circ| = k \geq 2$ , one can label its

nodes  $v_1 \leq \dots \leq v_k = v_*$ . Denoting the corresponding values of  $\eta$  by  $\eta_1, \dots, \eta_k$ , we see that in this case our assumptions are equivalent to the fact that, for each  $j \in \{1, \dots, k\}$ , one has  $\sum_{i \geq j} \eta_i > 0$  and, if  $j > 1$ ,  $\sum_{i < j} \eta_i < 0$ . In this case, we define  $\tilde{T}$  to be the tree where we remove the root  $v_1$  and take  $v_2$  as our new root. Similarly, to above, we define  $\tilde{\eta}$  on  $\tilde{T}^\circ$  by setting it equal to  $\eta$  except on  $v_2$  where we set  $\tilde{\eta}_{v_2} = \eta_{v_2} + \eta_{v_1}$ . We then have the bound

$$\begin{aligned} \mathcal{I}_\lambda(\eta) &= \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} \prod_{\omega \in T^\circ} 2^{-\ell_\omega \eta_\omega} = \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} \sum_{0 \leq m \leq \ell_{v_2}} 2^{-m \eta_{v_1}} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \eta_\omega} \\ &\sim \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} 2^{-\ell_{v_2} \eta_{v_1}} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \eta_\omega} = \sum_{\ell \in \mathcal{N}_\lambda(\tilde{T}^\circ)} \prod_{\omega \in \tilde{T}^\circ} 2^{-\ell_\omega \tilde{\eta}_\omega} = \mathcal{I}_\lambda(\tilde{\eta}), \end{aligned}$$

as above. Again, we note that  $\tilde{\eta}$  does again satisfy our assumptions, so the claim follows from the inductive hypothesis. Since we have exhausted all possibilities, this concludes the proof.  $\square$

**A.3. General form of the bound.** Given a labelled tree  $(T, \ell) \in \mathbb{T}(\mathbb{V})$ , denote by  $\mathcal{D}(T, \ell)$  the subset of  $(\mathbf{R}^d)^\mathbb{V}$  such that  $\|x_v - x_w\|_5 \leq |\mathbb{V}| 2^{-\ell_{v \wedge w}}$  for all  $v, w \in \mathbb{V}$ . As usual, we use the convention that  $x_0 = 0$ .

LEMMA A.13. Suppose  $(\tilde{K}^{(\mathbf{n})}(x))_{\mathbf{n}}$  is a family of functions such that for each  $\mathbf{n} \in \mathcal{N}(T, \ell)$ ,

$$\text{supp } \tilde{K}^{(\mathbf{n})} \subset \mathcal{D}(T, \ell) \quad (\text{A.17})$$

and

$$\int_{(\mathbf{R}^d)^\mathbb{V}} \hat{K}^{(\mathbf{n})}(x) dx = \int_{(\mathbf{R}^d)^\mathbb{V}} \tilde{K}^{(\mathbf{n})}(x) dx, \quad (\text{A.18})$$

then

$$|\mathcal{I}_\lambda(K)| \lesssim \sum_{(T, \ell) \in \mathbb{T}_\lambda(\mathbb{V})} \left( \prod_{v \in T^\circ} 2^{-\ell_v |\mathbb{S}|} \right) \sup_{\mathbf{n} \in \mathcal{N}(T, \ell)} \sup_x |\tilde{K}^{(\mathbf{n})}(x)|, \quad (\text{A.19})$$

where  $T^\circ$  denotes the set of interior nodes of  $T$ .

*Proof.* From (A.16), using (A.18) and (A.17),

$$|\mathcal{I}_\lambda(K)| \lesssim \sum_{T \in \mathbb{T}_\lambda(\mathbb{V})} \sum_{\mathbf{n} \in \mathcal{N}(T, \ell)} \left| \int_{\mathcal{D}(T, \ell)} \tilde{K}^{(\mathbf{n})}(x) dx \right|.$$



We claim that the Lebesgue measure of  $\mathcal{D}(T, \ell)$  is bounded from above by some fixed constant multiple of  $\prod_{v \in T^\circ} 2^{-\ell_v |s|}$ , which then immediately yields the required bound. To prove this, for each interior vertex  $v \in T^\circ$ , we choose two elements  $v_-, v_+ \in \mathbb{V}$  so that  $v_- \wedge v_+ = v$ . The collection of edges  $\{(v_-, v_+) : v \in T^\circ\}$  forms a spanning tree of  $\mathbb{V}$  and

$$\mathcal{D}(T, \ell) \subset \{x : \|x_{v_-} - x_{v_+}\|_s \leq |\mathbb{V}| 2^{-\ell_v} \quad \forall v \in T^\circ\}.$$

The claim follows by integrating over these coordinates one by one, from the leaves of the tree inwards.  $\square$

REMARK A.14. One could in principle simply choose  $\tilde{K}^{(n)} = \hat{K}^{(n)}$  in the lemma. It turns out that the resulting bound fails to take into account some cancellations and is not good enough for our purposes. The strategy of proof will then be to build, for each  $n \in \mathcal{N}(T, \ell)$ , a function  $\tilde{K}^{(n)}$  such that  $\sup_x |\tilde{K}^{(n)}(x)|$  can be bounded in a sharp way yielding a bound of the desired homogeneity in  $\lambda$ .

**A.4. Naive bound.** Define  $\eta : T^\circ \rightarrow \mathbf{R}$  by  $\eta(v) = |s| + \sum_{e \in \mathbb{E}} \eta_e(v)$ , where

$$\begin{aligned} \eta_e(v) = & -\hat{a}_e \mathbf{1}_{e_\uparrow}(v) + r_e (\mathbf{1}_{e_+ \wedge 0}(v) - \mathbf{1}_{e_\uparrow}(v)) \mathbf{1}_{r_e > 0, e_+ \wedge 0 > e_\uparrow} \\ & + (1 - r_e - \hat{a}_e) (\mathbf{1}_{e_- \wedge 0}(v) - \mathbf{1}_{e_\uparrow}(v)) \mathbf{1}_{r_e > 0, e_- \wedge 0 > e_\uparrow}, \end{aligned} \quad (\text{A.20})$$

with  $\mathbf{1}_v(v) = 1$  and  $\mathbf{1}_v(w) = 0$  for  $w \neq v$ , where we wrote  $e_\uparrow$  as a shorthand for  $e_- \wedge e_+$ . We then have the following bound for the functions  $\hat{K}^{(n)}$ :

LEMMA A.15. Assume that the  $\hat{K}^{(n)}$  are given by Definition A.5 with  $K_e$  satisfying (A.1) and  $\Psi^{(n)}$  given by (A.10). Then  $\eta$  defined by (A.20) is such that

$$\left( \prod_{v \in T^\circ} 2^{-\ell_v |s|} \right) \sup_x |\hat{K}^{(n)}(x)| \lesssim \prod_{v \in T^\circ} 2^{-\ell_v \eta(v)}, \quad (\text{A.21})$$

uniformly over all  $\mathbf{n} \in \mathcal{N}(T, \ell)$ .

*Proof.* Due to the multiplicative structure of both sides of this inequality, it holds as soon as, for any fixed edge  $e$ , we are able to prove the bound

$$\sup_x |\hat{K}_e^{(n_e)}(x_{e_-}, x_{e_+})| \lesssim \prod_{v \in T^\circ} 2^{-\ell_v \eta_e(v)}. \quad (\text{A.22})$$

Note here that the product on the right hand side actually only involves at most two terms as a consequence of (A.20). This bound in turn follows trivially from (A.9) for those edges  $e$  for which all  $r_e \leq 0$ .

For the edges with  $r_e > 0$ , and  $\mathbf{n}_e = (k, p, m)$  we will estimate in two different ways: if  $m > k$ , then we use [Hair14, Proposition A.1] to bound the next term in the Taylor expansion (A.11) (note if there were other multiedges  $\tilde{e} = (e_-, e_+)$  all have  $r_{\tilde{e}} = 0$  and so produce multiplicative factors  $K_{\tilde{e}}^{(k_{\tilde{e}})}(x_{e_+} - x_{e_-})$ ). This gives

$$\sup_x |\hat{K}_e^{(\mathbf{n}_e)}(x)| \lesssim 2^{-r_e m + (\hat{a}_e + r_e)p} \lesssim 2^{-r_e \ell_{e_+ \wedge 0} + (\hat{a}_e + r_e) \ell_{e_- \wedge 0}}, \quad (\text{A.23})$$

since  $\mathbf{n} \in \mathcal{N}(T, \ell)$  and therefore the index  $n_e = (k, p, m)$  satisfies  $|k - \ell_{e_\uparrow}| \leq c$ ,  $|p - \ell_{e_- \wedge 0}| \leq c$ , and  $|m - \ell_{e_+ \wedge 0}| \leq c$ . One checks that (A.22) holds for (A.20) using that in this case  $e_- \wedge 0 = e_\uparrow$ . On the other hand, if  $m \leq k$ , then we simply bound the terms appearing in (A.11) separately, which yields

$$\sup_x |\hat{K}_e^{(\mathbf{n}_e)}(x)| \lesssim 2^{\hat{a}_e k} + \sum_{|j|_S < r_e} 2^{-m|j|_S + (\hat{a}_e + |j|_S)p}. \quad (\text{A.24})$$

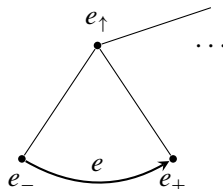
Then it is straightforward to check that (A.22) holds for (A.20). If  $e_- \wedge 0 = e_+ \wedge 0 < e_\uparrow$ , (A.24) gives a bound  $2^{\hat{a}_e k}$  as desired. If  $e_- \wedge 0 > e_\uparrow = e_+ \wedge 0$ , we have  $2^{\hat{a}_e k} \leq 2^{\hat{a}_e p}$  and  $p \geq m$  so  $2^{\hat{a}_e k} + \sum_{|j|_S < r_e} 2^{-m|j|_S + (\hat{a}_e + |j|_S)p} \lesssim 2^{(r_e - 1)(p - m) + \hat{a}_e p} \lesssim 2^{(r_e - 1)(p - k) + \hat{a}_e p}$ , as desired.  $\square$

The problem with this bound is that it is not the case in general that the function  $\eta$  satisfies the assumptions of Lemma A.10. This is because of the possible presence of edges  $e$  with  $\hat{a}_e > |s|$ , which can cause the first assumption of Lemma A.10 to fail. The purpose of the next subsection is to obtain an improved bound which deals with such a situation.

**A.5. Improved bound.** Let  $A^- \subset \mathbb{E}$  be the subset of those edges  $e$  such that the following two properties hold.

- One has  $r_e < 0$ .
- The element  $e_\uparrow \stackrel{\text{def}}{=} e_- \wedge e_+ \in T$  is such that if  $\{u, v\}$  are such that  $u \wedge v = e_\uparrow$ , then  $\{u, v\} = \{e_-, e_+\}$ .

In graphical terms, edges  $e \in A^-$  are those giving rise to the situation where the subtree of  $T$  below  $e_\uparrow$  consists only of the node  $e_\uparrow$  and the leaves  $e_-$  and  $e_+$ :



We now build a function  $\tilde{K}^{(\mathbf{n})}$  as follows. First, given any edge  $e = (e_-, e_+)$  and any  $r > 0$ , we define an operator  $\mathcal{D}_e^r$  acting on sufficiently smooth functions  $V: \mathbf{R}^{\mathbb{V}} \rightarrow \mathbf{R}$  by

$$(\mathcal{D}_e^r V)(x) = V(x) - \sum_{|k|_{\mathfrak{s}} < r} \frac{(x_{e_+} - x_{e_-})^k}{k!} (D_{e_+}^k V)(P_e(x)),$$

where  $D_{e_+}$  denotes differentiation with respect to the coordinate  $x_{e_+}$  and the function  $P_e: \mathbf{R}^{\mathbb{V}} \rightarrow \mathbf{R}^{\mathbb{V}}$  is given by

$$(P_e(x))_v = \begin{cases} x_v & \text{if } v \neq e_+, \\ x_{e_-} & \text{otherwise.} \end{cases}$$

We then further note that, as an immediate consequence of (A.12), the kernel  $\hat{K}^{(\mathbf{n})}$  factors naturally as

$$\hat{K}^{(\mathbf{n})}(x) = \hat{G}^{(\mathbf{n})}(x) \prod_{e \in A^-} \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}), \quad \hat{G}^{(\mathbf{n})}(x) = \prod_{e \notin A^-} \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}).$$

With these notations at hand, and writing  $A^- = \{e^{(1)}, \dots, e^{(k)}\}$  for some  $k \geq 0$ , we then define the kernel  $\tilde{K}^{(\mathbf{n})}$  by

$$\tilde{K}^{(\mathbf{n})}(x) = (\mathcal{D}_{e^{(k)}}^{r_{e^{(k)}}} \dots \mathcal{D}_{e^{(1)}}^{r_{e^{(1)}}} \hat{G}^{(\mathbf{n})})(x) \prod_{e \in A^-} \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+}). \quad (\text{A.25})$$

We can easily verify that one does indeed have the identity (A.18) because  $\hat{K}^{(\mathbf{n})}(x)$  and  $\tilde{K}^{(\mathbf{n})}(x)$  differ by a number of terms that are all of the form  $J(x)(x_{e_+} - x_{e_-})^k \hat{K}_e^{(\mathbf{n}_e)}(x_{e_-}, x_{e_+})$  where  $e \in A^-$ ,  $|k|_{\mathfrak{s}} < r_e$ , and where  $J$  is some smooth function depending on  $e$  and  $k$  that does not depend on the variable  $x_{e_+}$ . Integrating over  $x_{e_+}$  and using the fact that  $\hat{K}_e^{(\mathbf{n}_e)}$  annihilates polynomials of degree less than  $r_e$  by assumption, we conclude that (A.18) holds as claimed.

Now define  $\tilde{\eta}(v) = |\mathfrak{s}| + \sum_{e \in \hat{\mathbb{E}}} \tilde{\eta}_e(v)$  where

$$\tilde{\eta}_e(v) = \eta_e(v) + |r_e| \mathbf{1}_{e \in A^-} (\mathbf{1}_{e_{\uparrow}}(v) - \mathbf{1}_{e_{\uparrow}}(v)). \quad (\text{A.26})$$

Where  $\eta_e(v)$  is given in (A.20). Here  $e_{\uparrow} \in T^\circ$  denotes the ancestor of  $e_- \wedge e_+$ , that is, the element of the form  $w \wedge e_-$  with  $w \notin e$  which is furthest from the root. Note that there is at least one such  $w$  as long as  $e \in A^-$ , since either 0 or  $v_{\star,1}$  is a candidate. (If  $e \in A^-$  contains 0, it must be  $e_-$ , since  $e_+ \neq 0$  by the assumption that  $e_+ = 0$  implies  $r_e \geq 0$ . But then  $e_+ \neq v_{\star,1}$  since  $r_{(0, v_{\star,1})} = 0$  by assumption.)

LEMMA A.16. *The kernels  $\tilde{K}^{(\mathbf{n})}$  defined in (A.25) satisfy the bound*

$$\left( \prod_{v \in T^\circ} 2^{-\ell_v |\mathfrak{s}|} \right) \sup_x |\tilde{K}^{(\mathbf{n})}(x)| \lesssim \prod_{v \in T^\circ} 2^{-\ell_v \tilde{\eta}_v}, \quad (\text{A.27})$$

*uniformly over all  $\mathbf{n} \in \mathcal{N}(T, \ell)$ .*

REMARK A.17. Recalling Lemma A.10, and keeping in mind that the summation over labelled trees with vertex set  $\mathbb{V}$  can be absorbed into a  $(\mathbb{V}, \mathbb{E})$  dependent constant, we see that the proof of Theorem A.7 is complete as soon as we show that  $\tilde{\eta}$  does indeed satisfy the conditions of Lemma A.10, applied to the binary tree  $T^\circ$ , and is such that

$$|\tilde{\eta}| = |\mathfrak{s}||\mathbb{V}_0| - \sum_{e \in \hat{\mathbb{E}}} \hat{a}_e. \quad (\text{A.28})$$

*Proof.* Write  $\partial A^-$  for the set of all functions  $k: A^- \rightarrow \mathbf{N}^d$  such that, for all  $e \in A^-$ , one has  $|k_e|_{\mathfrak{s}} \geq |r_e|$  but  $|k_e - f|_{\mathfrak{s}} < |r_e|$  for some  $f \in \mathbf{N}^d$  with  $|f| = 1$ . For such a  $k$ , we write  $\mathbf{D}^k$  for the differential operator on  $(\mathbf{R}^d)^{A^-}$  given by  $\mathbf{D}^k = \prod_{e \in A^-} D_{x_{e+}}^{k_e}$ . With these notations at hand, it then follows from the construction of  $\tilde{K}^{(\mathbf{n})}$  and the generalized Taylor's formula [Hai14, Proposition A.1] that there are explicitly described positive measures  $\mathcal{Q}_x^{k,e}$  on  $\mathbf{R}^d$  with

$$\mathcal{Q}_x^{k,e}(\mathbf{R}^d) \lesssim \|x_{e+} - x_{e-}\|_{\mathfrak{s}}^{|k_e|_{\mathfrak{s}}}, \quad (\text{A.29})$$

such that one has the identity

$$\tilde{K}^{(\mathbf{n})}(x) = \left( \prod_{e \in A^-} \hat{K}_e^{(\mathbf{n}_e)}(x_{e-}, x_{e+}) \right) \sum_{k \in \partial A^-} \int_{(\mathbf{R}^d)^{A^-}} \mathbf{D}^k \hat{G}^{(\mathbf{n})}(x|y) \prod_{e \in A^-} \mathcal{Q}_x^{k,e}(dy_e), \quad (\text{A.30})$$

where we introduced the notation  $x|y$  for the element in  $(\mathbf{R}^d)^{\mathbb{V}_0}$  which is obtained by setting

$$(x|y)_v = \begin{cases} y_e & \text{if there is } e \in A^- \text{ such that } v = e-, \\ x_v & \text{otherwise.} \end{cases}$$

This definition makes sense thanks to our assumption that there are not multiple edges  $e \in A^-$  emerging from the same vertex, and because we are using smooth approximations to the distributional kernels.

Furthermore, it follows similarly to before that if  $r_e \leq 0$  then, for every such multiindex  $k$ , one has the bound

$$\sup_x |D_{e\pm}^k \hat{K}_e^{(\mathbf{n}_e)}(x)| \lesssim 2^{\ell_{e+}|k|_{\mathfrak{s}}} \prod_{v \in T^\circ} 2^{-\ell_v \eta_e(v)},$$

uniformly over  $n \in \mathcal{N}_c(T, \ell)$ . If  $r_e > 0$  on the other hand, one obtains the bound

$$\sup_x |D_{e\pm}^k \hat{K}_e^{(\mathbf{n}_e)}(x)| \lesssim (2^{\ell_{e+}|k|_{\mathfrak{s}}} + 2^{\ell_{e\pm \wedge 0}|k|_{\mathfrak{s}}}) \prod_{v \in T^\circ} 2^{-\ell_v \eta_e(v)}.$$

Combining this with the bound (A.29), the definition of  $\eta$ , and the fact that one has  $|k_e|_{\mathfrak{s}} \geq |r_e|$  for every edge  $e \in A^-$ , we conclude that the function  $\tilde{K}^{(\mathbf{n})}$  satisfies

$$\sup_x |\tilde{K}^{(\mathbf{n})}(x)| \lesssim \left( \prod_{v \in T^\circ} 2^{-\ell_v \eta(v)} \right) \left( \prod_{e \in A^-} 2^{(\ell_{e_\uparrow} - \ell_{e_\downarrow})|r_e|} \right), \quad (\text{A.31})$$

which is precisely the required bound.  $\square$

REMARK A.18. By the definition of the set of edges  $A^-$ , for every  $e \in A^-$  and every  $w \notin e$ , one always has the property that  $e_\pm \wedge w < e_\uparrow$ , so that the exponent appearing in the second factor above is always negative. In other words, our choice of the set  $A^-$  guarantees that the bound (A.31) is always an improvement over (A.21).

**A.6. Putting everything together.** By Remark A.17 the following lemma, which is the final statement of this section, completes the proof of Theorem A.7.

LEMMA A.19. *The function  $\tilde{\eta}$  given in (A.26) satisfies the identity (A.28) and the assumptions of Lemma A.10 (applied to the tree  $T$ ).*

*Proof.* To verify that assumption 1 of Lemma A.10 holds, we choose an arbitrary element  $v \in T^\circ$  and we consider the set  $L_v \subset \mathbb{V}$  of all the leaves  $u \in T$  with  $u \geq v$ . Note that one always has  $|L_v| \geq 2$ , and we will treat the case  $|L_v| = 2$  separately.

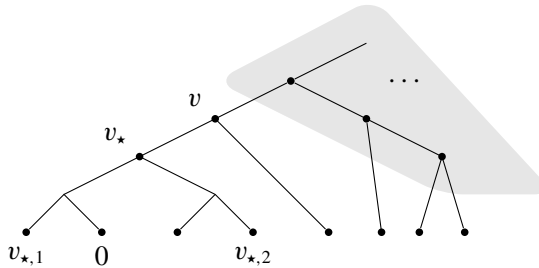
If  $|L_v| = 2$ , then there exists an edge  $e$  such that  $L_v = e$  and  $v = e_\uparrow$ . In this case, assumption 1 of Lemma A.10 requires that  $\tilde{\eta}(v) > 0$ . We have  $\tilde{\eta}(v) = |\mathfrak{s}| - \hat{a}_e + |r_e| \mathbf{1}_{r_e < 0}$  which is indeed positive by Assumption A.1(1).

We now turn to the case  $|L_v| > 2$ . Since  $e_\uparrow > e_\uparrow^\dagger$ , one always has  $\sum_{u \geq v} (\mathbf{1}_{e_\uparrow}(u) - \mathbf{1}_{e_\uparrow^\dagger}(u)) \geq 0$ . From the definitions (A.20) of  $\eta$  and (A.26) of  $\tilde{\eta}$  we have  $\sum_{u \geq v} \tilde{\eta}(u) \geq \sum_{u \geq v} \eta(u)$ . By checking all cases  $e_- \in L_v$ ,  $e_- \notin L_v$ ,  $e_+ \in L_v$ ,  $e_+ \notin L_v$ ,  $0 \in L_v$ ,  $0 \notin L_v$  separately, we find that  $\sum_{e \in \hat{\mathbb{E}}} \sum_{u \geq v} \eta_e(u)$  is given by

$$\sum_{e \in \hat{\mathbb{E}}_0} -\hat{a}_e + \mathbf{1}_{0 \in L_v} \left( \sum_{e \in \hat{\mathbb{E}}_+^\downarrow} r_e + \sum_{e \in \hat{\mathbb{E}}_+^\uparrow} (-\hat{a}_e - r_e + 1) \right),$$

with  $\hat{\mathbb{E}} = \hat{\mathbb{E}}(L_v)$ ,  $\mathbb{E}_0 = \mathbb{E}_0(L_v)$ , and so on. Note the cancellation which appears in the special case when all three  $e_-$ ,  $e_+$ ,  $0 \in L_v$ . Now points 2 and 3 of Assumption A.1 with the choice  $\tilde{\mathbb{V}} = L_v$  imply that  $\sum_{u \geq v} \eta(u) > 0$ , which concludes the proof that assumption 1 holds.

We now turn to the second condition appearing in Lemma A.10. In our case, we choose for the distinguished node  $v_\star$  the most recent common ancestor between the elements of  $\mathbb{V}_\star$ . The reason for this choice is that this node encodes the largest scale appearing in the multiscale clustering which is still guaranteed to be smaller than the scale  $\lambda$  fixed by the test function. We then fix an arbitrary node  $v \in T^\circ$  such that  $v_\star \geq v$ . Denoting by  $U_v = \{u \in T^\circ : u \not\geq v\}$ , the situation is the following, where  $U_v$  contains all the nodes lying in the shaded region:



Note again that similarly to before, one has  $\sum_{u \in U_v} \tilde{\eta}(u) \leq \sum_{u \in U_v} \eta(u)$ , so that we can restrict ourselves to the verification of the second condition for the function  $\eta$ . Denoting by  $\bar{\mathbb{V}}$  the set of leaves attached to  $U_v$ , one has  $\bar{\mathbb{V}} \subset \mathbb{V} \setminus \mathbb{V}_\star$ . By checking the three cases directly we have

$$\sum_e \sum_{u \in U_v} \eta_e(u) = - \sum_{e \in \hat{\mathbb{E}} \setminus \hat{\mathbb{E}}_+^\downarrow} \hat{a}_e - \sum_{e \in \hat{\mathbb{E}}_+^\uparrow} r_e + \sum_{e \in \hat{\mathbb{E}}_+^\downarrow} (r_e - 1)$$

with the obvious notation that  $\hat{\mathbb{E}} = \hat{\mathbb{E}}(\bar{\mathbb{V}})$ ,  $\hat{\mathbb{E}}_+^\uparrow = \hat{\mathbb{E}}_+^\uparrow(\bar{\mathbb{V}})$ , and so on. Furthermore, the cardinality of  $U_v$  is exactly equal to  $|\bar{\mathbb{V}}|$  so we have

$$\sum_{u \in U_v} \eta(u) = |\mathfrak{s}| |\bar{\mathbb{V}}| - \sum_{e \in \hat{\mathbb{E}} \setminus \hat{\mathbb{E}}_+^\downarrow} \hat{a}_e - \sum_{e \in \hat{\mathbb{E}}_+^\uparrow} r_e + \sum_{e \in \hat{\mathbb{E}}_+^\downarrow} (r_e - 1),$$

so that the condition  $\sum_{u \in U_v} \eta(u) < 0$  is satisfied as a consequence of Assumption A.1(4).

Finally to see that it satisfies the identity (A.28), note that similar to before, termwise cancellation gives us  $\sum_{v \in T} \tilde{\eta}_e(v) = -\hat{a}_e$ . Hence  $|\tilde{\eta}| = \sum_{v \in T^\circ} (|\mathfrak{s}| + \sum_{e \in \hat{\mathbb{E}}} \tilde{\eta}_e(v)) = |\mathfrak{s}| |T^\circ| - \sum_{e \in \hat{\mathbb{E}}} \hat{a}_e$ . Since  $T^\circ$  is a binary tree,  $|T^\circ| = \# \text{ of leaves} - 1 = |\mathbb{V}_0|$ .  $\square$

## Appendix B. Notes on renormalization

Recall that given a map  $M: \mathcal{T}_{\text{ex}} \rightarrow \mathcal{T}_{\text{ex}}$  as in Section 5.2 and the associated maps  $\hat{M}$  and  $\Delta^M$  defined by (5.7), the map  $\hat{\Delta}^M: \mathcal{T}_+ \rightarrow \mathcal{T}_+ \otimes \mathcal{T}_+$  appearing in

[**Hai14**, Equation 8.38] is uniquely determined by the relation

$$(\mathcal{A}\hat{M}\mathcal{A} \otimes \hat{M})\Delta^+ = (1 \otimes \mathcal{M})(\Delta^+ \otimes 1)\hat{\Delta}^M, \quad (\text{B.1})$$

where  $\mathcal{A}: \mathcal{T}_+ \rightarrow \mathcal{T}_+$  is the antipode of the Hopf algebra  $\mathcal{T}_+$  defined as in [**Hai14**, Theorem 8.16] and  $\Delta^+$  is its coproduct given in [**Hai14**, Equation 8.9]. The motivation for the definition (**B.1**) is that if  $(\Pi, \Gamma)$  is an admissible model and one defines  $(\Pi^M, \Gamma^M)$  by

$$\Pi_x^M = (\Pi_x \otimes f_x)\Delta^M, \quad \gamma_{xy}^M = (\gamma_{xy} \otimes f_y)\hat{\Delta}^M, \quad (\text{B.2})$$

(with  $\Gamma_{xy}\tau = (1 \otimes \gamma_{xy})\Delta\tau$  and similarly for  $\Gamma_{xy}^M$ ) then  $(\Pi^M, \Gamma^M)$  does satisfy all the algebraic identities required for an admissible model.

In [**Hai14**], the renormalization group associated to a regularity structure generated by noises, products and abstract integration maps was defined as the set of maps  $M$  preserving the noises and  $\mathbf{1}$ , commuting with the abstract integration maps and multiplication by  $X^k$ , such that furthermore both  $\Delta^M$  and  $\hat{\Delta}^M$  are ‘upper triangular’ in the sense that, if we write

$$\Delta^M \tau = \tau^{(1)} \otimes \tau^{(2)}, \quad \hat{\Delta}^M \bar{\tau} = \bar{\tau}^{(1)} \otimes \bar{\tau}^{(2)},$$

with an implicit summation suppressed in the notation, then one has  $|\tau^{(1)}| \geq |\tau|$  and  $|\bar{\tau}^{(1)}| \geq |\bar{\tau}|$ . This property was absolutely crucial since this is what guarantees that if we use  $\Delta^M$  and  $\hat{\Delta}^M$  to renormalize a model as in (**B.2**), then  $(\Pi^M, \Gamma^M)$  also satisfies the *analytical* properties required to be a model. In this section, we show that one only ever needs to verify that  $\Delta^M$  is upper triangular, as this then automatically implies the same for  $\hat{\Delta}^M$ .

Throughout this section, we consider a general regularity structure generated by a number of ‘noise symbols’  $\mathcal{E}$ , a multiplication operation, as well as a number of abstract integration operators. In other words, every basis vector of  $\mathcal{T}$  is assumed to be generated from the vectors  $\mathcal{E}_i$ ,  $X_i$  or  $\mathbf{1}$  by multiplication and/or abstract integration. The structure considered in this article is of this type since  $\mathcal{E}^k$  can be considered as an integration operator of order  $|k|$ . Our main result can be summarized as follows.

**THEOREM B.1.** Let  $(\mathcal{T}, \mathcal{G})$  be a regularity structure as above and let  $M: \mathcal{T} \rightarrow \mathcal{T}$  be a linear map preserving  $\mathcal{E}_i$ ,  $X^k$ , and commuting with the abstract integration maps and with multiplication by  $X^k$ . Let  $\Delta^M$  and  $\hat{\Delta}^M$  be given by (**B.1**). If  $\Delta^M$  is upper triangular, then so is  $\hat{\Delta}^M$ .

In order to prove Theorem **B.1**, we first derive a number of identities involving the operators  $\Delta^M$  and  $\hat{\Delta}^M$ . We first note that a simple calculation using the

coassociativity of  $\Delta^+$  and the properties of the antipode  $\mathcal{A}$  yields

$$((1 \otimes \mathcal{M})(\Delta^+ \otimes 1))^{-1} = (1 \otimes \mathcal{M})(1 \otimes \mathcal{A} \otimes 1)(\Delta^+ \otimes 1),$$

so that the first identity can be rewritten somewhat more explicitly as

$$\hat{\Delta}^M = (1 \otimes \mathcal{M})((1 \otimes \mathcal{A})\Delta^+ \hat{\mathcal{M}} \mathcal{A} \otimes \hat{M})\Delta^+. \quad (\text{B.3})$$

Similarly, the second identity is equivalent to

$$\Delta^M = (1 \otimes \mathcal{M})((1 \otimes \mathcal{A})\Delta M \otimes \hat{M})\Delta. \quad (\text{B.4})$$

Throughout this section, we will make use of the following notation. Given a map  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, k\}$ , we write  $\mathcal{M}^\sigma$  for the map

$$\mathcal{M}^\sigma : \bigotimes_{j=1}^n \tau_j \mapsto \bigotimes_{i=1}^k \left( \prod_{j \in \sigma^{-1}(i)} \tau_j \right).$$

For a surjection  $\sigma$ , we also use the notation  $\sigma = (\sigma^{-1}(1)) \cdots (\sigma^{-1}(k))$ , so that for example

$$\mathcal{M}^{(2)(1,4)(2,5)}(\tau_1 \otimes \cdots \otimes \tau_5) = \tau_2 \otimes (\tau_1 \tau_4) \otimes (\tau_2 \tau_5).$$

It will also sometimes be convenient to use for the above example the alternative notation

$$\mathcal{M}^{(2)(1,4)(2,5)} = \mathcal{M}^{(2)} \otimes \mathcal{M}^{(1,4)} \otimes \mathcal{M}^{(2,5)}.$$

Recall also that the antipode  $\mathcal{A}$  is automatically an antihomomorphism of coalgebras [Swe69], so that

$$\Delta^+ \mathcal{A} = \mathcal{M}^{(2)(1)}(\mathcal{A} \otimes \mathcal{A})\Delta^+.$$

With all of these notations at hand, we first claim that one has the following.

LEMMA B.2. *The identity*

$$\mathcal{M}^{(1,3)(2)}(\hat{M} \otimes \hat{\Delta}^M \mathcal{A})\Delta^+ = (1 \otimes \mathcal{A})\Delta^+ \hat{M}, \quad (\text{B.5})$$

holds true.

*Proof.* In view of (B.2), it is natural to test both sides of (B.5) against  $f_y \otimes \gamma_{xy}$ . Since  $\gamma_{xy} = f_x^{-1} \circ f_y$ , the right hand side is then equal to  $(f_y \circ \gamma_{xy}^{-1})\hat{M} = f_x \hat{M} = f_x^M$ . The left hand side on the other hand is equal to

$$(f_y \otimes \gamma_{xy} \otimes f_y)(\hat{M} \otimes \hat{\Delta}^M \mathcal{A})\Delta^+ = (f_y^M \otimes (\gamma_{xy}^M)^{-1})\Delta^+ = f_y^M \circ (\gamma_{xy}^M)^{-1} = f_x^M,$$



as required since  $f_x$  and  $f_y$  are arbitrary multiplicative functionals. More directly, it follows from (B.3) that

$$\begin{aligned} \mathcal{M}^{(1,3)(2)}(\hat{M} \otimes \hat{\Delta}^M \mathcal{A}) \Delta^+ &= \mathcal{M}^{(1,3,4)(2)}(\hat{M} \otimes ((1 \otimes \mathcal{A}) \Delta^+ \mathcal{A} \hat{M} \mathcal{A} \otimes \hat{M}) \Delta^+ \mathcal{A}) \Delta^+ \\ &= \mathcal{M}^{(1,2,4)(3)}(\hat{M} \otimes (\hat{M} \mathcal{A} \otimes (1 \otimes \mathcal{A}) \Delta^+ \mathcal{A} \hat{M}) \Delta^+) \Delta^+ \\ &= \mathcal{M}^{(1,2,4)(3)}((\hat{M} \otimes \hat{M} \mathcal{A}) \Delta^+ \otimes (1 \otimes \mathcal{A}) \Delta^+ \mathcal{A} \hat{M}) \Delta^+ \\ &= \mathcal{M}^{(2)(1)}(1 \otimes \mathcal{A}) \Delta^+ \mathcal{A} \hat{M} = (1 \otimes \mathcal{A}) \Delta^+ \hat{M}, \end{aligned}$$

as required.  $\square$

LEMMA B.3. *One has the identity*

$$((1 \otimes \mathcal{A}) \Delta \otimes 1) \Delta^M = \mathcal{M}^{(1)(3,4)(2,5)}((1 \otimes \Delta^+) \Delta^M \otimes (\mathcal{A} \otimes 1) \hat{\Delta}^M) \Delta.$$

*Proof.* It follows from the definitions of  $\Delta^M$  and  $\hat{\Delta}^M$  that the right hand side is given by

$$\begin{aligned} \mathcal{M}^{(1)(3,4)(2,5)}((1 \otimes \Delta^+) \Delta^M \otimes (\mathcal{A} \otimes 1) \hat{\Delta}^M) \Delta &= \mathcal{M}^{(1)(3,4)(2,5,6)}((1 \otimes \Delta^+ \mathcal{M})((1 \otimes \mathcal{A}) \Delta M \otimes \hat{M}) \Delta \\ &\quad \otimes ((\mathcal{A} \otimes \mathcal{A}) \Delta^+ \mathcal{A} \hat{M} \mathcal{A} \otimes \hat{M}) \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3,5)(2,4,6)}((1 \otimes \Delta^+ \mathcal{M})((1 \otimes \mathcal{A}) \Delta M \otimes \hat{M}) \Delta \otimes (\Delta^+ \hat{M} \mathcal{A} \otimes \hat{M}) \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3,5,7)(2,4,6,8)}(((1 \otimes \Delta^+ \mathcal{A}) \Delta M \otimes \Delta^+ \hat{M}) \Delta \otimes (\Delta^+ \hat{M} \mathcal{A} \otimes \hat{M}) \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3,5,7)(2,4,6,8)}((1 \otimes \Delta^+ \mathcal{A}) \Delta M \otimes \Delta^+ \hat{M} \otimes \Delta^+ \hat{M} \mathcal{A} \otimes \hat{M}) (\Delta \otimes \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3,5,7)(2,4,6,8)}((1 \otimes \Delta^+ \mathcal{A}) \Delta M \otimes (\Delta^+ \hat{M} \otimes \Delta^+ \hat{M} \mathcal{A}) \Delta^+ \otimes \hat{M}) (1 \otimes \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3,5)(2,4,6)}((1 \otimes \Delta^+ \mathcal{A}) \Delta M \otimes \Delta^+ \hat{M} \mathcal{M} (1 \otimes \mathcal{A}) \Delta^+ \otimes \hat{M}) (1 \otimes \Delta^+) \Delta \\ &= \mathcal{M}^{(1)(3)(2,4)}((1 \otimes \Delta^+ \mathcal{A}) \Delta M \otimes \hat{M}) \Delta \\ &= (1 \otimes 1 \otimes \mathcal{M})((1 \otimes (\mathcal{A} \otimes \mathcal{A}) \Delta^+) \Delta M \otimes \hat{M}) \Delta \\ &= (1 \otimes 1 \otimes \mathcal{M})(((1 \otimes \mathcal{A}) \Delta \otimes \mathcal{A}) \Delta M \otimes \hat{M}) \Delta \\ &= ((1 \otimes \mathcal{A}) \Delta \otimes 1) (1 \otimes \mathcal{M}) ((1 \otimes \mathcal{A}) \Delta M \otimes \hat{M}) \Delta = ((1 \otimes \mathcal{A}) \Delta \otimes 1) \Delta^M \end{aligned}$$

as required.  $\square$

LEMMA B.4. *One has the identity*

$$\mathcal{I}_k(\tau) \otimes \mathbf{1} - \mathbf{1} \otimes \mathcal{I}_k(\tau) = \sum_{\ell} \left( 1 \otimes \frac{(-X)^\ell}{\ell!} \mathcal{M} \right) ((1 \otimes \mathcal{A}) \Delta^+ \mathcal{I}_{k+\ell} \otimes 1) \Delta \tau.$$

(B.6)

*Proof.* It follows from the recursive definition of  $\Delta^+$  that

$$\begin{aligned} & \sum_{\ell} \left( 1 \otimes \frac{(-X)^{\ell}}{\ell!} \mathcal{M} \right) ((1 \otimes \mathcal{A}) \Delta^+ \mathcal{J}_{k+\ell} \otimes 1) \Delta \tau \\ &= \sum_{\ell, m} \left( 1 \otimes \frac{(-X)^{\ell}}{\ell!} \mathcal{M} \right) \left( \left( \mathcal{J}_{k+\ell+m} \otimes \frac{X^m}{m!} \mathcal{A} \right) \Delta \otimes 1 \right) \Delta \tau \\ &+ \sum_{\ell} 1 \otimes \frac{(-X)^{\ell}}{\ell!} \mathcal{M} (\mathcal{A} \mathcal{J}_{k+\ell} \otimes 1) \Delta \tau. \end{aligned}$$

It now follows from the defining property of  $\mathcal{A}$ , followed by the binomial identity and the comodule property of  $\Delta$  and  $\Delta^+$  (see the statement and proof of [Hai14, Theorem 8.16]) that

$$\begin{aligned} & \sum_{\ell} \frac{(-X)^{\ell}}{\ell!} \mathcal{M} (\mathcal{A} \mathcal{J}_{k+\ell} \otimes 1) \Delta \tau \\ &= - \sum_{\ell, m} \frac{(-X)^{\ell}}{\ell!} \mathcal{M} \left( \mathcal{M} \left( \mathcal{J}_{k+\ell+m} \otimes \frac{X^m}{m!} \mathcal{A} \right) \Delta \otimes 1 \right) \Delta \tau \\ &= - \mathcal{M} (\mathcal{M} (\mathcal{J}_k \otimes \mathcal{A}) \Delta \otimes 1) \Delta \tau \\ &= - \mathcal{M} (\mathcal{M} (\mathcal{J}_k \otimes \mathcal{A}) \otimes 1) (1 \otimes \Delta^+) \Delta \tau \\ &= - \mathcal{M} (\mathcal{J}_k \otimes \mathcal{M} (\mathcal{A} \otimes 1) \Delta^+) \Delta \tau = - \mathcal{J}_k(\tau). \end{aligned}$$

Here, we used the fact that  $\mathcal{M}(\mathcal{A} \otimes 1) \Delta^+ = \mathbf{1}^*$  and  $(1 \otimes \mathbf{1}^*) \Delta \tau = \tau$ . Similarly, it follows from the binomial identity followed by the comodule property that

$$\begin{aligned} & \sum_{\ell, m} \left( 1 \otimes \frac{(-X)^{\ell}}{\ell!} \mathcal{M} \right) \left( \left( \mathcal{J}_{k+\ell+m} \otimes \frac{X^m}{m!} \mathcal{A} \right) \Delta \otimes 1 \right) \Delta \tau \\ &= (1 \otimes \mathcal{M}) ((\mathcal{J}_k \otimes \mathcal{A}) \Delta \otimes 1) \Delta \tau \\ &= (\mathcal{J}_k \otimes \mathcal{M} (\mathcal{A} \otimes 1) \Delta^+) \Delta \tau = \mathcal{J}_k(\tau) \otimes \mathbf{1}, \end{aligned}$$

which concludes the proof of (B.6).  $\square$

In the sequel, we will write  $\mathcal{Q}_{>\alpha}$  for the projection onto elements of homogeneity greater than  $\alpha - 2$  (the number 2 being the gain of homogeneity given by the integration operator  $\mathcal{J}$ ). We will also denote by  $\mathcal{Q}_{\geq\alpha}$  the projection onto elements of homogeneity at least  $\alpha - 2$ , so that  $\mathcal{Q}_{\geq\alpha} + \mathcal{Q}_{<\alpha} = I$ . We now have all the ingredients required to obtain a recursive characterization of  $\hat{\Delta}^M$  from which Theorem B.1 can then easily be derived.

PROPOSITION B.5. The map  $\hat{\Delta}^M$  satisfies the identity

$$\begin{aligned} \hat{\Delta}^M \mathcal{J}_k(\tau) &= (\mathcal{J}_k \otimes 1) \Delta^M \tau - \sum_{\ell} \left( \frac{X^{\ell}}{\ell!} \mathcal{M}^{(2,3)} \otimes \mathcal{M}^{(1,4)} \right) \\ &\quad \times ((1 \otimes \mathcal{A}) \Delta^+ \mathcal{M}(\mathcal{J}_{k+\ell} \otimes 1) \Delta^M \mathcal{Q}_{\leq |k+\ell|} \otimes \hat{\Delta}^M) \Delta \tau. \end{aligned} \quad (\text{B.7})$$

*Proof.* We apply the ‘swapping’ operator  $\mathcal{M}^{(2)(1)}: \tau \otimes \bar{\tau} \mapsto \bar{\tau} \otimes \tau$  to (B.6) and then apply the map  $(1 \otimes \mathcal{M})(\hat{\Delta}^M \otimes \hat{M})$  to both sides. This yields the identity

$$\begin{aligned} \hat{\Delta}^M \mathcal{J}_k \tau &= \mathbf{1} \otimes \hat{M} \mathcal{J}_k \tau \\ &\quad + \sum_{\ell} \left( \frac{(-X)^{\ell}}{\ell!} \otimes \mathcal{M} \right) (\hat{\Delta}^M \mathcal{M}^{(2,3)} \otimes \mathcal{M}^{(1)}) \\ &\quad \times ((\hat{M} \otimes \mathcal{A}) \Delta^+ \mathcal{J}_{k+\ell} \otimes 1) \Delta \tau \\ &= \mathbf{1} \otimes \hat{M} \mathcal{J}_k \tau \\ &\quad + \sum_{\ell} \left( \frac{(-X)^{\ell}}{\ell!} \mathcal{M}^{(2,4)} \otimes \mathcal{M}^{(1,3,5)} \right) ((\hat{M} \otimes \hat{\Delta}^M \mathcal{A}) \Delta^+ \mathcal{J}_{k+\ell} \otimes \hat{\Delta}^M) \Delta \tau \\ &= \mathbf{1} \otimes \hat{M} \mathcal{J}_k \tau \\ &\quad + \sum_{\ell} \left( \frac{(-X)^{\ell}}{\ell!} \mathcal{M}^{(2,3)} \otimes \mathcal{M}^{(1,4)} \right) ((1 \otimes \mathcal{A}) \Delta^+ \hat{M} \mathcal{J}_{k+\ell} \otimes \hat{\Delta}^M) \Delta \tau, \end{aligned} \quad (\text{B.8})$$

where we made use of Lemma B.2 to obtain the last identity. We furthermore use the fact that

$$\hat{M} \mathcal{J}_k = \mathcal{M}(\mathcal{J}_k \otimes 1) \Delta^M \mathcal{Q}_{>|k|},$$

by (5.7) (we can insert the operator  $\mathcal{Q}_{>|k|}$  since  $\mathcal{J}_k = \mathcal{J}_k \mathcal{Q}_{>|k|}$ ), which leads to the identity

$$\begin{aligned} &((1 \otimes \mathcal{A}) \Delta^+ \hat{M} \mathcal{J}_{k+\ell} \otimes \hat{\Delta}^M) \Delta \tau \\ &= ((1 \otimes \mathcal{A}) \Delta^+ \mathcal{M}(\mathcal{J}_{k+\ell} \otimes 1) \Delta^M \mathcal{Q}_{>|k+\ell|} \otimes \hat{\Delta}^M) \Delta \tau. \end{aligned} \quad (\text{B.9})$$

Noting that

$$\mathbf{1} \otimes \hat{M} \mathcal{J}_k(\tau) = \mathbf{1} \otimes \mathcal{M}(\mathcal{J}_k \otimes 1) \Delta^M \tau, \quad (\text{B.10})$$

we then make use of Lemma B.4 which yields

$$\begin{aligned} \mathbf{1} \otimes \mathcal{M}(\mathcal{J}_k \otimes 1) \Delta^M \tau &= (\mathcal{J}_k \otimes 1) \Delta^M \tau - \sum_{\ell} \left( \frac{(-X)^{\ell}}{\ell!} \mathcal{M}^{(2,3)} \otimes \mathcal{M}^{(1,4)} \right) \\ &\quad \times (((1 \otimes \mathcal{A}) \Delta^+ \mathcal{J}_{k+\ell} \otimes 1) \Delta \otimes 1) \Delta^M \tau. \end{aligned} \quad (\text{B.11})$$

(Here we used the fact that the left hand side, and therefore also the right hand side, of (B.6) is symmetric under the map  $\tau \otimes \bar{\tau} \mapsto \bar{\tau} \otimes \tau$ .) At this stage we use the fact that, thanks to Lemma B.3, one has

$$\mathcal{M}^{(1)(3,4)(2,5)}((1 \otimes (1 \otimes \mathcal{A})\Delta^+) \Delta^M \otimes \hat{\Delta}^M) \Delta = (\Delta \otimes 1) \Delta^M.$$

(just compose both sides with  $1 \otimes \mathcal{A} \otimes 1$ ), which then yields the identity

$$\begin{aligned} & \mathcal{M}^{(2,3)(1,4)}((1 \otimes \mathcal{A})\Delta^+ \mathcal{M}(\mathcal{J}_k \otimes 1) \Delta^M \otimes \hat{\Delta}^M) \Delta \\ &= \mathcal{M}^{(2,4,5)(1,3,6)}(((1 \otimes \mathcal{A})\Delta^+ \mathcal{J}_k \otimes (1 \otimes \mathcal{A})\Delta^+) \Delta^M \otimes \hat{\Delta}^M) \Delta \\ &= \mathcal{M}^{(2,3)(1,4)}((1 \otimes \mathcal{A})\Delta^+ \mathcal{J}_k \otimes 1 \otimes 1) \\ &\quad \times \mathcal{M}^{(1)(3,4)(2,5)}((1 \otimes (1 \otimes \mathcal{A})\Delta^+) \Delta^M \otimes \hat{\Delta}^M) \Delta \\ &= \mathcal{M}^{(2,3)(1,4)}(((1 \otimes \mathcal{A})\Delta^+ \mathcal{J}_k \otimes 1) \Delta \otimes 1) \Delta^M. \end{aligned}$$

Combining this with (B.11), (B.10), (B.9) and (B.8) finally leads to the required identity.  $\square$

We can now finally turn to the proof of Theorem B.1.

*Proof of Theorem B.1.* We proceed by induction. Assume that the statement holds for all the elements in  $\mathcal{T}_+$  appearing in the description of  $\Delta\tau$ , then we claim that the statement also holds for  $\mathcal{J}_k(\tau)$  as well as for  $\mathcal{E}_k^\ell(\tau)$ . Since the algebraic properties of  $\mathcal{E}_k^\ell$  are the same as those of  $\mathcal{J}_k$ , we only consider the latter. For the first term in (B.7), this follows from the upper triangular structure of  $\Delta^M$ . Regarding the second term, it follows from the induction hypothesis that the quantity

$$(\mathcal{Q}_{<|k+\ell|} \otimes \hat{\Delta}^M) \Delta\tau,$$

is necessarily a linear combination of expressions of the form  $\tau^{(1)} \otimes \tau^{(2)} \otimes \tau^{(3)}$  with  $|\tau^{(1)}| + |\tau^{(2)}| \geq |\tau|$  and  $|\tau^{(1)}| < |k + \ell| - 2$ . In particular, one has  $|\tau^{(2)}| \geq |\tau| + 2 - |k + \ell|$ . It now suffices to note that, with  $\tau^{(i)}$  as just defined for any fixed  $\ell$ , the second term in (B.7) always consists of linear combinations of terms of the form

$$X^\ell \tau^{(2)} \sigma^{(1)} \otimes \sigma^{(2)} \tau^{(3)},$$

with the  $\tau^{(i)}$  as above and some  $\sigma^{(i)}$  in  $\mathcal{T}_+$ . Since the  $\sigma^{(i)}$  belong to  $\mathcal{T}_+$ , they have positive homogeneity, so that the homogeneity of the first factor is at least  $|\tau^{(2)}| + |\ell| = |\tau| + 2 - |k + \ell| + |\ell| = |\mathcal{J}_k(\tau)|$ , thus concluding the proof.  $\square$

## Appendix C. Symbolic index

In this appendix, we collect some of the most used symbols of the article, together with their meaning and the page where they were first introduced.

Symbol	Meaning	Page
$ z  = \ z\ _s$	Parabolic distance	13
$ k  =  k _s$	Parabolic length of a multiindex	19
$ \tau $	Homogeneity of $\tau \in \mathcal{T}$	17
$ \tau _\alpha$	(Euclidean) norm of projection of $\tau$ onto $\mathcal{T}_\alpha$	26
$\mathcal{B}$	Smooth compactly supported functions $\varphi: \mathbf{R}^2 \rightarrow \mathbf{R}$	23
$C^W$	Wick renormalization constant	48
$\Delta$	Linear map $\mathcal{T} \rightarrow \mathcal{T} \otimes \mathcal{T}_+$ used to build structure group	19
$\mathcal{D}$	Abstract spatial derivative	20
$\mathcal{D}^{\gamma, \eta}, \ \cdot\ _{\gamma, \eta}$	Modelled distributions with blow-up at $t = 0$	27
$\mathcal{D}_\varepsilon^{\gamma, \eta}, \ \cdot\ _{\gamma, \eta; \varepsilon}$	Modelled distributions with short range $\varepsilon$	36
$\mathcal{E}^k$	Abstract symbol of multiplication by $\varepsilon^k$ on $\mathcal{T}_{\text{ex}}$	18
$\mathcal{E}_\ell^k(\tau)$	Formal symbol representing $\varepsilon^k D^{(\ell)} \Pi_z \tau(z)$	19
$\hat{\mathcal{E}}^k$	Abstract symbol on $\mathcal{D}^\gamma$	29
$F$	Nonlinearity	4
$\mathcal{G}$	Structure group	19
$\mathcal{G}_+$	Multiplicative linear functionals on $\mathcal{T}_+$	20
$\gamma_{z\bar{z}}$	Character defining $\Gamma_{z\bar{z}}$	28
$\ H\ _\gamma, \ H; \bar{H}\ _\gamma$	(Hölder) norm on $\mathcal{D}_\gamma$	26
$H_k$	$k$ th Hermite polynomial	49
$\mathcal{I}, \mathcal{I}'$	Abstract integration maps	17
$\mathcal{I}_\ell(\tau)$	Formal symbol representing $D^{(\ell)} K \star \Pi_z \tau(z)$	19
$\mathcal{K}$	Operator on $\mathcal{D}^\gamma$ defined from kernel $K$	32
$\mathcal{L}_\varepsilon(\zeta)$	Lift of smooth $\zeta$ to $\mathcal{T}_{\text{ex}}$	27
$\mathcal{M}$	Admissible models	23
$\mathcal{M}_\varepsilon$	$\varepsilon$ -models	35
$\mathcal{M}_{\text{ex}}$	Admissible models on $\mathcal{T}_{\text{ex}}$	26
$M^{\text{Wick}}$	Wick renormalization	49
$N_\varepsilon$	Kernel appearing in the model bounds	67
$\mathcal{P}$	Operator defined from the heat kernel	32
$ \Pi ,  \Pi; \bar{\Pi} $	Norm on models	25

Symbol	Meaning	Page
$Q_\varepsilon^{(m)}$	Kernel appearing in the model bounds	67
$Q_{\leq 0}$	Projection onto $\bigoplus_{\alpha \leq 0} \mathcal{T}_\alpha$ in $\mathcal{T}_{\text{ex}}$	31
$\mathcal{R}$	Reconstruction operator	26
$\mathcal{R}^{\text{Wick}}$	Reconstruction operator for Wick renormalized model	54
$\mathcal{R}$	Kernel renormalization	87
$\mathcal{S}'$	Dual of Schwartz space	24
$\mathcal{T}$	Linear span of $\mathcal{W}$	17
$\mathcal{T}_{<\gamma}$	Elements of $\mathcal{T}$ of homogeneity strictly less than $\gamma$	27
$\tilde{\mathcal{T}}$	Subspace of $\mathcal{T}$ generated by powers of $X$	19
$\mathcal{T}_+$	Free commutative algebra generated by $X$ , $\mathcal{I}_\ell(\tau)$ , $\mathcal{O}_\ell^k(\tau)$	19
$\mathcal{T}_{\text{ex}}$	Extended regularity structure	21
$\mathcal{U}$	Symbols used to describe solution $h$	17
$\mathcal{U}'$	Symbols used to describe derivative $\partial_x h$	17
$\mathcal{U}'_{\text{ex}}$	Symbols used to describe powers of $\partial_x h$	21
$\mathcal{V}$	Symbols used to describe right hand side	17
$\mathcal{V}_{\ell,k}$	Symbols that can appear as argument to $\mathcal{O}_\ell^k$	19
$\mathcal{W}$	Symbols used to describe equation	17
$\mathcal{W}_{\text{ex}}$	Symbols used for the extended regularity structure	22
$\mathcal{W}_+$	Symbols used to describe structure group	19
$\bar{\mathcal{W}}_+$		22
$\mathcal{E}$	Abstract symbol for noise	17
$X = (X_0, X_1)$	Abstract symbol for $z = (t, x)$	17
$\xi$	Space–time white noise	2

## References

- [ACQ11] G. Amir, I. Corwin and J. Quastel, ‘Probability distribution of the free energy of the continuum directed random polymer in 1 + 1 dimensions’, *Comm. Pure Appl. Math.* **64**(4) (2011), 466–537.
- [AKQ10] T. Alberts, K. Khanin and J. Quastel, ‘Intermediate disorder regime for directed polymers in dimension 1+1’, *Phys. Rev. Lett.* **105**(9) (2010), 090603.
- [BA07] M. Ben-Artzi, ‘Lectures on viscous Hamilton-Jacobi equations’, unpublished lecture notes (2007). URL: <http://www.ma.huji.ac.il/~mbartzi/>.
- [BC14] A. Borodin and I. Corwin, ‘Macdonald processes’, *Probab. Theory Related Fields* **158**(1–2) (2014), 225–400.

- [BG97] L. Bertini and G. Giacomin, ‘Stochastic Burgers and KPZ equations from particle systems’, *Comm. Math. Phys.* **183**(3) (1997), 571–607.
- [BP57] N. N. Bogoliubow and O. S. Parasiuk, ‘Über die Multiplikation der Kausalfunktionen in der Quantentheorie der Felder’, *Acta Math.* **97**(1957) 227–266.
- [BPRS93] L. Bertini, E. Presutti, B. Rüdiger and E. Saada, ‘Dynamical fluctuations at the critical point: convergence to a nonlinear stochastic PDE’, *Teor. Veroyatnost. i Primenen.* **38**(4) (1993), 689–741.
- [BQS11] M. Balázs, J. Quastel and T. Seppäläinen, ‘Fluctuation exponent of the KPZ / stochastic Burgers equation’, *J. Amer. Math. Soc.* **24**(3) (2011), 683–708.
- [CT15] I. Corwin and L.-C. Tsai, ‘KPZ equation limit of higher-spin exclusion processes’, *Ann. Probab.* **45**(3) (2017), 1771–1798.
- [DT13] A. Dembo and L.-C. Tsai, ‘Weakly asymmetric non-simple exclusion process and the Kardar-Parisi-Zhang equation’, *Comm. Math. Phys.* **341**(1) (2016), 219–261.
- [FG17] M. Furlan and M. Gubinelli, ‘Weak universality for a class of 3d stochastic reaction-diffusion models’, Preprint, 2017, [arXiv:1708.03118](https://arxiv.org/abs/1708.03118).
- [FH14] P. Friz and M. Hairer, *A Course on Rough Paths*, Universitext (Springer, New York, 2014), xiv+251. With an introduction to Regularity Structures. doi:[10.1007/978-3-319-08332-2](https://doi.org/10.1007/978-3-319-08332-2).
- [FQ14] T. Funaki and J. Quastel, ‘KPZ equation, its renormalization and invariant measures’, *Stoch. Partial Differ. Equ. Anal. Comput.* **3**(2) (2015), 159–220.
- [GJ14] P. Gonçalves and M. Jara, ‘Nonlinear fluctuations of weakly asymmetric interacting particle systems’, *Arch. Ration. Mech. Anal.* **212**(2) (2014), 597–644.
- [GP15] M. Gubinelli and N. Perkowski, ‘KPZ reloaded’, *Comm. Math. Phys.* **349**(1) (2017), 165–269.
- [GRZ17] Y. Gu, L. Ryzhik and O. Zeitouni, ‘The Edwards-Wilkinson limit of the random heat equation in dimensions three and higher’, Preprint, 2017, [arXiv:1710.00344](https://arxiv.org/abs/1710.00344).
- [Hai13] M. Hairer, ‘Solving the KPZ equation’, *Ann. of Math.* (2) **178**(2) (2013), 559–664.
- [Hai14] M. Hairer, ‘A theory of regularity structures’, *Invent. Math.* **198**(2) (2014), 269–504.
- [Hai15a] M. Hairer, ‘Introduction to regularity structures’, *Braz. J. Probab. Stat.* **29**(2) (2015), 175–210.
- [Hai15b] M. Hairer, ‘Regularity structures and the dynamical  $\Phi_3^4$  model’, in *Current Developments in Mathematics 2014* (Int. Press, Somerville, MA, 2016), 1–49.
- [Hep69] K. Hepp, ‘On the equivalence of additive and analytic renormalization’, *Comm. Math. Phys.* **14**(1969) 67–69.
- [HHZ95] T. Halpin-Healy and Y.-C. Zhang, ‘Kinetic roughening phenomena, stochastic growth, directed polymers and all that’, *Phys. Rep.* **254**(4–6) (1995), 215–414.
- [HL15] M. Hairer and C. Labbé, ‘Multiplicative stochastic heat equations on the whole space’, *J. Eur. Math. Soc. (JEMS)* **20**(4) (2018), 1005–1054.
- [HP15] M. Hairer and É. Pardoux, ‘A Wong-Zakai theorem for stochastic PDEs’, *J. Math. Soc. Japan* **67**(4) (2015), 1551–1604.
- [HPP13] M. Hairer, É. Pardoux and A. Piatnitski, ‘Random homogenization of a highly oscillatory singular potential’, *SPDEs: Anal. Comput.* **1**(4) (2013), 571–605.
- [HS15] M. Hairer and H. Shen, ‘A central limit theorem for the KPZ equation’, *Ann. Probab.* **45**(6B) (2017), 4167–4221.
- [HX18] M. Hairer and W. Xu, ‘Large-scale limit of interface fluctuation models’, Preprint, 2018, [arXiv:1802.08192](https://arxiv.org/abs/1802.08192).
- [KPZ86] M. Kardar, G. Parisi and Y.-C. Zhang, ‘Dynamic scaling of growing interfaces’, *Phys. Rev. Lett.* **56**(9) (1986), 889–892.

- [Kru56] J. B. Kruskal Jr., ‘On the shortest spanning subtree of a graph and the traveling salesman problem’, *Proc. Amer. Math. Soc.* **7**(1956) 48–50.
- [KS91] J. Krug and H. Spohn, ‘Kinetic roughening of growing surfaces’, in *Solids Far from Equilibrium* (Cambridge University Press, Cambridge, 1991), 479–583.
- [Mey92] Y. Meyer, ‘Wavelets and operators’, in *Cambridge Studies in Advanced Mathematics* vol. 37 (Cambridge University Press, Cambridge, 1992), xvi+224. Translated from the 1990 French original by D. H. Salinger. doi:[10.1017/CBO9780511623820](https://doi.org/10.1017/CBO9780511623820).
- [MFQR] G. Moreno-Flores, J. Quastel and D. Remenik, In preparation.
- [MU17] J. Magnen and J. Unterberger, ‘Diffusive limit for 3-dimensional KPZ equation: the Cole-Hopf case’, Preprint, 2017, [arXiv:1702.03122](https://arxiv.org/abs/1702.03122).
- [Mue91] C. Mueller, ‘On the support of solutions to the heat equation with noise’, *Stochastics and Stochastic Reports* **37**(4) (1991), 225–245.
- [Nua06] D. Nualart, ‘The Malliavin calculus and related topics’, in *Probability and its Applications (New York)*, second edn, (Springer, Berlin, 2006), xiv+382.
- [PP12] É. Pardoux and A. Piatnitski, ‘Homogenization of a singular random one-dimensional PDE with time-varying coefficients’, *Ann. Probab.* **40**(3) (2012), 1316–1356.
- [QV13] J. Quastel and B. Valkó, ‘Diffusivity of lattice gases’, *Arch. Ration. Mech. Anal.* **210**(1) (2013), 269–320.
- [Spo91] H. Spohn, *Large Scale Dynamics of Interacting Particles*, Texts and Monographs in Physics (Springer, Berlin, 1991).
- [Swe69] M. E. Sweedler, ‘Hopf algebras’, in *Mathematics Lecture Note Series* (W. A. Benjamin, Inc., New York, 1969), vii+336.
- [Wal86] J. B. Walsh, ‘An introduction to stochastic partial differential equations’, in *École d’été de probabilités de Saint-Flour, XIV—1984*, Lecture Notes in Mathematics, 1180 (Springer, Berlin, 1986), 265–439.
- [Zim69] W. Zimmermann, ‘Convergence of Bogoliubov’s method of renormalization in momentum space’, *Comm. Math. Phys.* **15**(1969) 208–234.